Solution NMR studies of HIV-1 reverse transcriptase

by

Naima Gabriela Sharaf

Bachelor of Science in Chemistry, University of North Carolina-Chapel Hill, 2008

Submitted to the Graduate Faculty of the

School of Medicine in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2016

UNIVERSITY OF PITTSBURGH

SCHOOL OF MEDICINE

This dissertation was presented

by

Naima Gabriela Sharaf

It was defended on

November 21, 2016

and approved by

Dr. Rieko Ishima, Associate Professor, Department of Structural Biology

Dr. Patrick C.A van der Wel, Assistant Professor, Department of Structural Biology

Dr. Ronald C. Montelaro, Professor, Department of Microbiology and Molecular Genetics

Dissertation advisor: Dr. Angela Gronenborn, Professor, Department of structural Biology

SOLUTION NMR STUDIES OF HIV-1 REVERSE TRANSCRIPTASE

Naima Gabriela Sharaf, PhD

University of Pittsburgh, 2016

The human immunodeficiency virus (HIV-1) is a well-recognized threat to global public health, with current estimates of ~37 million infected individuals worldwide [1]. While efforts to develop an effective vaccine to prevent infection have not come to fruition, several FDA approved drugs targeting HIV-1 have proven effective in reducing viral replication and improving quality of life [2]. Yet, due to the increasing emergence of drug resistant variants, new HIV-1 treatments are needed as the virus continues to evolve. Structure-based drug design is a powerful approach for facilitating the development of new HIV-1 inhibitors [3], driven mainly by x-ray crystallographic studies of viral proteins [4]. However, the static nature of crystallographic studies necessitate complementary structural and dynamic studies to genuinely understand protein-ligand interactions. In this thesis, solution NMR is used to investigate HIV-1 RT, a major therapeutic target in the fight against HIV-1 infection. HIV-1 reverse transcriptase produces viral DNA from genomic RNA and is essential for the viral lifecycle [5]. Although the structure of the mature heterodimer has been well characterized, less is known about RT maturation from its p66 immature precursor, which so far has resisted crystallization. I used NMR to investigate RT maturation and to elucidate RT interactions with FDA approved inhibitors. Together the results of my thesis contribute to arrive at a more comprehensive understanding of RT maturation and RT-inhibitor interactions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I would like to dedicate this work to my two beautiful sons,

Julius William Barnes and Maximus Alexander Barnes.

I would like to express my deep gratitude to my mentor Dr. Angela Gronenborn for her guidance, support, and encouragement. I would also like to thank Drs. Rieko Ishima and In-Ja Byeon for their valuable insight and advice. Special thanks should be given to Mike Delk for NMR technical support and the members of the Gronenborn and Ishima laboratories for insightful discussions. I would like to specifically thank Dr. Matthew Whitley, for insightful discussions and critical reading of parts of this thesis. I would also like to thank the University of Pittsburgh whose family friendly policies significantly contributed to my success as a graduate student. Special thanks goes to the University Child Development Center and the Pittsburgh public schools for providing exceptional childcare. I would also like to recognize my best friend, Christopher Barnes, for your love, help and support. Last, special recognition goes out to my family for always believing in me.

ABBREVIATIONS

HIV-1	Human Immunodeficiency virus
FDA	Food and Drug Administration
RT	HIV-1 reverse transcriptase
AIDS	Acquired Immunodeficiency Syndrome
TF	transframe protein
CA	capsid protein
IN	integrase
PR	protease
AZT	zidovudine
NRTI	nucleoside analog RT inhibitor
ddI	didanosine
ddC	zalcitabine
d4T	stavudine
3TC	lamivudine
PI	protease inhibitors
SQV	saquinavir
IDV	indinavir

NNRTI	non-nucleoside RT inhibitors
HAART	Highly Active Antiretroviral Therapy
FI	fusion inhibitor
EI	entry inhibitor
INSTs	integrase inhibitor
RNH	RNase H
NOE	Nuclear Overhauser effects
NOESY	Nuclear Overhauser Enhancement Spectroscopy
DD	dipole-dipole relaxations
CSA	chemical-shift anisotropy

1.0 INTRODUCTION

1.1 HIV-1

HIV-1 targets cells in the human immune system, increasing an individual's susceptibility to infections and some types of cancers [1]. The most advanced phase of HIV-1 infection is referred to as Acquired Immunodeficiency Syndrome (AIDS). As of 2014, the World Health Organization estimates that HIV-1 has claimed over 34 million lives, and ~37 million people are currently infected with HIV-1. Therefore, the HIV-1 pandemic continues to be a global health issue [1]. In response, over the past 25 years, numerous multidisciplinary studies on the HIV-1 virus have been carried out, resulting in a large breadth of knowledge about the virus. HIV-1 belongs to the *Retroviridae* viral family, which comprises enveloped viruses containing RNA genomes [1]. HIV-1 targets cells expressing the CD4 receptor and the chemokine receptors CCR5 or CXCR4, including macrophages and CD4+ T cells (**Figure 1.1a)** [6]. The fusion of viral and host cell membranes is mediated by the interaction between the viral Env glycoprotein complex and the host cell CD4 receptors [7].

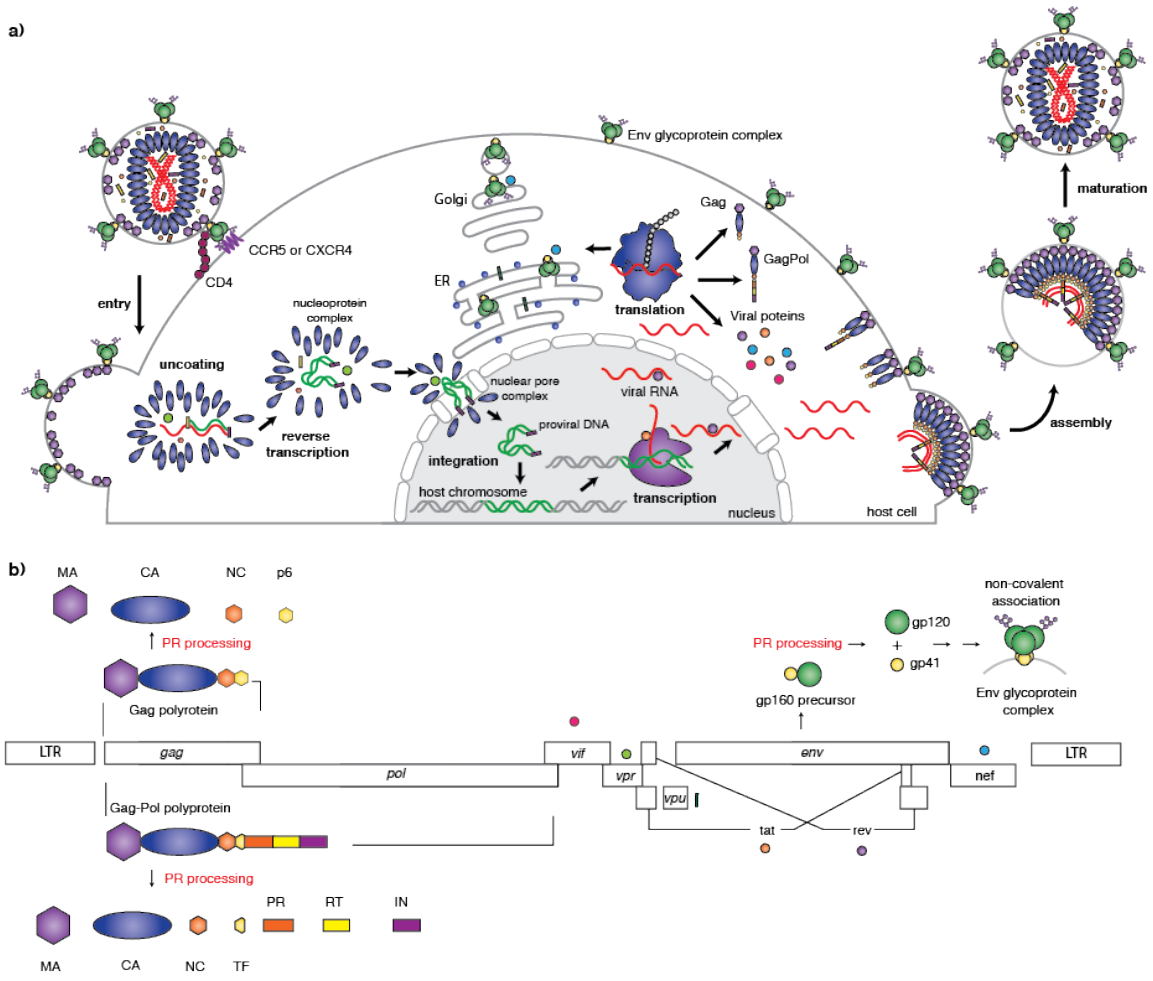

Figure 1.1 Schematic diagram of the HIV-1 life cycle and the organization of the HIV-1 genome

a) HIV-1 viral replication can not occur in isolation. Therefore, the HIV-1 virus relies on the ability to hijack a host cell's molecular machinery to replicate. The HIV-1 life cycle begins with the fusion of the viral and host membrane and the subsequent release of the capsid into the host cell's cytoplasm. Once inside the cell, the single stranded RNA viral genome is converted into double stranded proviral DNA and then shuttled into the nucleus. After the proviral DNA is integrated into the host chromosome, the host's molecular machinery produces numerous copies of viral RNA. Viral RNA transcripts can be either translated into viral proteins or directly packaged into progeny viruses. After packaging and viral detachment, PR cleaves Gag and Gag-Pol polyproteins, causing morphological changes to produce a mature virus. The newly produced virus can then infect neighboring cells and produce more progeny

viruses by reiterating through the HIV-1 life cycle (figure inspired by [8] and [9]). b) Schematic representation of the proteins coded by the HIV-1 genome (adapted from [10]).

Once the viral and host membranes are fused, the cone-shaped capsid core is released into the host cell cytoplasm. The capsid is composed of capsid protein (CA), which oligomerizes into hexamers and a few pentamers, [11] and surrounds the viral components, including the mature enzymes integrase (IN), protease (PR) and reverse transcriptase (RT), and the RNA genome. While still in the cytosol, RT produces double stranded DNA using genomic viral RNA as the template. Then, the viral DNA associates with other proteins to form a nucleoprotein complex, which is subsequently shuttled through the nuclear pore with the aid of Vpr, a viral accessory protein with no enzymatic activity[11]. Inside the nucleus, the viral genome is integrated into the host chromosome by IN-mediated DNA processing and strand transfer, in conjunction with host cell enzymes that are hijacked for the subsequent steps in DNA integration. Once the viral DNA is covalently inserted into the host's chromosome, the host's molecular machinery produces numerous copies of the viral RNA that can be directly packaged into progeny viruses or translated into viral proteins [12].

In the nucleus, the viral RNA transcripts are processed by splicing to generate various protein-encoding mRNAs. The structural proteins, Gag and Gag-Pol are produced from unspliced mRNA, while the viral proteins proteins Vif, Vpu, Vpr, and Env are produced from singly spliced mRNA, and Tat, Rev and Nef are produced from multi-spliced mRNA [6, 12, 13]. All viral proteins have distinct roles in the lifecycle. Tat upregulates viral RNA transcription by interacting with host cell elongations factors and RNA polymerase II (RNAP II), increasing RNAP II processivity and, consequently, the production of viral RNA by ~100 fold. Rev helps facilitate the transport of the singly spliced and unspliced viral RNA transcripts out of the

nucleus to the cytoplasm for translation and packaging [14]. Vif protects progeny viral DNA from damage by mediating the degradation of apolipoprotein B mRNA-editing enzyme-catalytic polypeptide-like 3G (APOBEC3G), a host immune surveillance protein that introduces hyper-mutations into the viral DNA by catalyzing the deamination of deoxycytidine to deoxyuridine [10, 15, 16].

The structural proteins Env, Gag and Gag-Pol also play distinct roles in the viral lifecycle. The *env* gene codes for the gp160 precursor, which is cleaved to produce the Env proteins gp120 and gp41. The Env proteins are shuttled to the host cell plasma membrane through the secretory pathway, via the rough endoplasmic reticulum and golgi. In mature virons, gp120 and gp41 proteins associate non-covalently to form trimeric structures, which sparsely stud the viral envelope. The Env glycoprotein complex is the main target for host cell antibodies; however, heavy glycosylation conceals surface epitopes, allowing the virus to evade host cell immune proteins [17].

The Gag polyprotein is encoded by *gag* gene, while the Gag-Pol polyprotein is produced as a result of a frame shift event during translation that joins the *gag* and *pol* reading frames (**Figure 1.1b**). Together, the intact Gag and Gag-Pol polyproteins play an essential role shuttling and packaging viral components, including viral RNA and viral proteins, into the progeny virus. Once the progeny virus detaches, both Gag and Gag-Pol polyproteins are processed by PR. Gag polyprotein is then cleaved to produce the proteins MA, CA, NC, p6 and two spacer proteins p1 and p2. Since the Gag and Gag-Pol polyproteins share the same N-terminal sequence, cleavage of the Gag-Pol polyprotein also produces MA, CA, and NC. In addition, PR cleavage of the Gag-Pol polyprotein also produces the transframe protein (TF) and the subunits essential in the formation of mature RT, IN, and RT. As a result of PR processing, the virus undergoes

morphological changes called maturation, which is marked by the formation of the fullerene cone. The detached mature progeny virus can then infect neighboring cells and reiterate through the viral lifecycle to produce more progeny viruses [8, 18].

1.1.1 HIV-1 treatment and drug resistance

Although new evidence suggests HIV-1 spread to the US in the early 1970, awareness and recognition of HIV-1 as a serious health condition only began in 1981 with the first description of a previously healthy patient with a rare opportunistic infection. By the end of 1981, over 200 similar cases had been reported [8, 19-23]. Due to lack treatment options at this time, patient care often focused on the management of opportunistic infections and AIDS related symptoms. In 1987, HIV-1 specific treatment was revolutionized when the Food and Drug Administration (FDA) approved zidovudine (AZT), a nucleoside analog RT inhibitor (NRTI) lacking a 3′ hydroxyl that prevents the effective production of proviral DNA by causing early chain termination during reverse transcription. By 1996, five other NRTIs were approved including didanosine (ddI), zalcitabine (ddC), stavudine (d4T), and lamivudine (3TC) [24]. Initially NRTIs were given as mono-drug therapy and dual-drug therapy, which produced positive treatment outcomes including the increased life-span of infected individuals. But, these benefits were temporary, and treatment with NRTIs led to the rapid development of resistant HIV-1 variants [25].

Subsequent advances in HIV-1 treatment were made with the development of other classes of drugs. Protease inhibitors (PI) saquinavir (SQV) and indinavir (IDV) were approved in 1995 and 1996, respectively, which bind to the active site of PR and act as competitive inhibitors. In 1996 and 1998, the FDA approved nevirapine (NVP) and efavirenz (EFV),

respectively, which inhibit RT allosterically by binding to a pocket distinct from the active site. Due to their distinct mechanism of action, these inhibitors were named non-nucleoside RT inhibitors (NNRTIs). In the late 1990s, several important studies demonstrated the benefits of triple-drug therapy, which typically comprise drugs from three different drug classes, including NRTIs, NNRTIs, and PIs [26-29]. This treatment was named Highly Active Antiretroviral Therapy (HAART) and became the standard of care by 1998. With proper adherence, this treatment regimen has produced declines in the rates of hospitalization, AIDS, and death [3, 30]. Currently, 24 inhibitors are approved by the FDA. These include second- and third- generation RT and PR inhibitors, including NRTIs, NNRTIs, and PIs. In addition, a fusion inhibitor (FI), an entry inhibitor (EI) and integrase inhibitors (INSTs) have also been developed. Many of these FDA approved drugs were discovered using high-throughput methods. Of note, the development of some of these drugs, in particular the second- and third-generation inhibitors, can be attributed to significant advances in structural biology, which facilitated the structure determination of many HIV-1 enzymes and led to the success of structure-based drug design [3]. As a result of decades worth of research, HIV-1 infection has changed from a life-threatening disease to a manageable chronic condition.

However, even with current treatments, HIV-1 infection can lead to the emergence of drug resistant HIV-1 variants [2]. The emergence of drug resistant HIV-1 variants can be attributed to the virus's high genetic variability, which is facilitated by an error prone RT that does not correct errors by proofreading. Consequently, RT has an error rate per detectable nucleotide of 1/1700, compared to 1/17000 and 1/30000 in avian and myeloblastosis and murine leukemia viruses, respectively. In addition, mutational hotspots were found where the error rate was higher, a much as 1/70 [31]. As a result, of the high genetic variability and short life cycle

(half-life of ~2 days), almost complete replacement of wild-type (wt) virus in plasma by drug-resistant variants can occur after fourteen days [32, 33].

Interestingly, it has been found that a single transmitted/founder virus establishes a productive HIV-1 infection. But, over time, the high genetic variability of the viruses leads to the production of HIV-1 variants [34, 35]. As a result, within an infected individual there can be a heterogeneous mixture of HIV-1 variants that compete for resources necessary for replication. The relative population of each HIV-1 variant relies on many factors including its adaptation to the host's intracellular environment. As a result, in the presence of HIV-1 inhibitors, only HIV-1 variants with reduced susceptibility will persist, shifting the relative population of HIV-1 variants and eventually leading to therapy failure. Currently, resistance has been documented for all antiretroviral drug classes. Therefore new HIV-1 treatments will be needed as the virus continues to evolve and acquire drug resistance mutations [2].

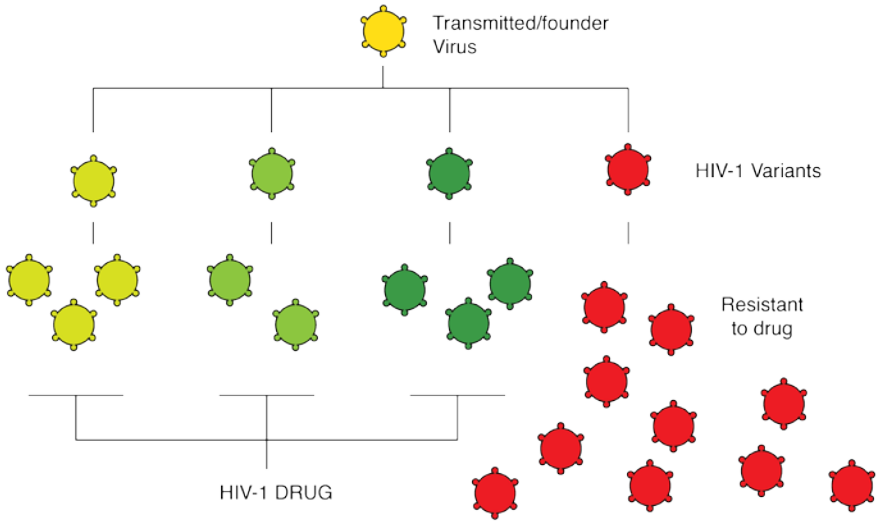

Figure 1.2 HIV-1 variants within an infected individual

a) Although HIV-1 infection begins with a single transmitted/founder virus, overtime a heterogeneous mixture of HIV-1 variants can arise within an infected individual, all compete for resources. As a result, HIV-1 variants with reduced susceptibility to HIV-1 drugs will persist leading to treatment failure.

1.1.2 HIV-1 RT

Since RT is essential for HIV-1 replication, RT continues to be an attractive target for therapeutic interventions, [4, 36]. To develop novel inhibitors with more favorable resistance profiles, it is important to gain a thorough understanding of RT, its structure, and mechanism of inhibitor action. Towards this goal, significant efforts have been made to gain structural insights into RT, drug resistant variants of RT, and RT-inhibitor interactions.

X-ray crystallography has been invaluable to obtain the molecular details of RT. There are over 200 of crystal structures of RT currently available [37-57]. A comparison of the available X-ray crystal structures shows that mature HIV-1 RT is an asymmetric heterodimer composed of two subunits named p66 and p51. The p66 subunit contains two domains, the

polymerase and RNase H (RNH) domains. The p66 subunit is further divided into subdomains, named fingers, thumb, palm, and connection. The arrangements of the subdomains in the p66 subunit resemble a right hand. The p51 subunit shares the same N-terminal sequence as the p66 subunit, but lacks the RNH domain. Consequently, the p51 subunit also contains the fingers, thumb, palm and connection subdomains but, given the asymmetry of the mature enzyme, the subdomains have different spatial arrangements within the subunit (**Figure 1.3).**

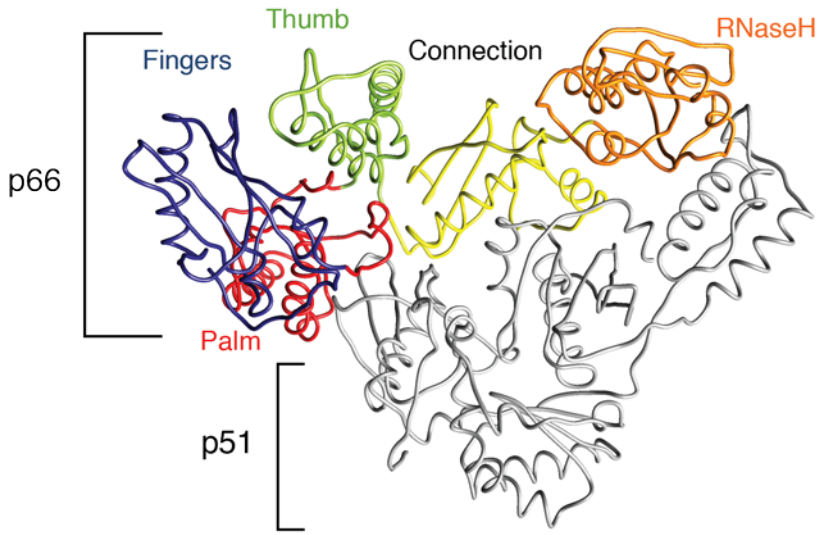

Figure 1.3 General description of RT structure

Tube representation of apo-RT (PDB: 1DLO [52]), with the fingers, palm, thumb, connection, and RNH domains in the p66 subunit colored in blue, red, green, yellow and orange, respectively. The p51 subunit is colored grey.

Structures representing several functional states of RT have been reported including: apo-RT, DNA-RT complex, and dNTP-DNA-RT complex. Analysis of these X-ray structures reveals major conformational rearrangements upon substrate binding, resulting from hinge movements of the subdomains relative to each other, while the structure of individual subdomains remains

almost invariant. The defining feature of most apo-RT X-ray crystal structures is the “closed conformation” of the fingers and thumb subdomains in the p66 subunit, where the fingers and thumb subdomain appear to clamp down on each other [38]. In contrast, in DNA-RT crystal structures the fingers and thumb subdomains in the p66 subunit are in the “open conformation”, forming a large cleft to bind dsDNA substrate [39]. The dsDNA stretches from the polymerase active site to the RNH active site interacting with the connection subdomains of both the p66 and p51 subunits, which is ~60 Å in length. The binding of dNTP results in the fingers subdomain folding down in the presence of dsDNA, closer to the palm subdomain [37-39].

X-ray structures of RT in the presence of several inhibitors are also available. A comparison of crystal structures of apo-RT and RT in the presence of NNRTIs, shows significant structural changes upon NNRTI binding. In the presence of NNRTIs, p66 thumb adopts an “open conformation”, where the thumb is 30 Å away from the fingers subdomain. Changes are also observed in the NNRTI-binding pocket, located in the p66 subunit, which is lined by aromatic (Y181, Y188, F227, W229, and Y232), hydrophobic (P59, L100, V106, V179, L234, and P236), and hydrophilic (K101, K103, S105, D132, and E224) residues [52]. Most NNRTIs contain a common pharmacophore scaffold, including moieties capable of hydrogen bonding, aromatic ring(s) that participate in π-π interactions, and hydrophobic domain able to interact with the hydrophobic region of the NNRTI-binding pocket [36, 46, 58, 59]. X-ray crystallography has been invaluable for providing details of enzyme-substrate interaction. However, studies by other methods are necessary to provide true understanding of the systems under investigation. Here we use solution NMR to study RT in solution.

1.2 NMR SPECTROSCOPY OF PROTEINS

1.2.1 A brief history of protein NMR

The pioneering work of many researchers laid the foundation for modern protein NMR spectroscopy, including early studies by Wolfgang Pauli in 1924, who proposed that nuclei should posses spin angular momentum [60], and Isidor Isaac Rabi, who demonstrated the nuclear resonance effect from metal ions in a vacuum [61]. In 1945 groups led by Felix Bloc and Edward Purcell demonstrated resonance effects from solid and liquid samples. Early protein NMR studies including the first ^{1}H NMR spectrum of a protein ribonuclease A in 1957 by Martin Saunders [62] and a spectral comparison between native and denatured protein by Bovery in 1959 [63, 64].

These early studies were often difficult due low sensitivity and the narrow chemical shift dispersion in the proton dimension [65]. Therefore, subsequent advances in instrumentation and processing methodologies aided development of NMR studies on biological macromolecules. Some of the important developments include the introduction of signal averaging [66, 67] and pulsed Fourier transform NMR [68], which reduced acquisition time and quantity of protein time required. In addition, significant advances were made in the development of high field magnets, and by 1952 spectrometers became commercially available [65]. In 1967, McDonald and Phillips reported the first use of a superconducting magnet (220 MHz) to obtain NMR spectra of protein. Their work demonstrated that improved spectra could be acquired at higher field strengths [69]. This work helped stimulate interest in the development of even more powerful magnets.

Following this time, there was significant progress brought about by the work of several laboratories. In the 1970s Richard Ernst was a key figure in the introduction of two-dimensional

(2D) NMR experiments, which provide information on correlations between two nuclei (either via J connectivities or cross relaxation pathways), and result in spectra with two frequency dimensions [70]. Shortly after, Kurt Wüthrich applied 2D correlation spectroscopy to study proteins, and by 1983 his group was routinely using a variety of ^{1}H NMR experiments to obtain the resonance assignments of several small proteins [71, 72]. The first solution structure of a protein, which was calculated using 2D NOE distance restraints, was published in 1985 Kurt Wüthrich [73]. By the second half of the 1980s, ^{1}H NMR became an established tool for the structure determination of proteins up to 100 residues [74].

In the late 1980s, structure calculations of proteins larger than 100 residues were still difficult. One challenge for large proteins is the spectral overlap caused by the high number of signals, making it difficult to resolve and unambiguously assign individual resonances. To address this problem, 3D NMR experiments were introduced [74-81], first with unlabeled proteins, and then with ^{13}C and ^{15}N labeled proteins [80, 82]. With these new sets of experiments, resonance assignments were considerably easier, allowing the determination of larger protein structures. Also in the late 1980s, more robust methods for structure determinations were developed by Clore and Gronenborn [75, 78, 79]. With their simulated annealing methods, which combine experimentally determined restraints with molecular dynamic simulations, NMR structures could be calculated with comparable precision and accuracy of 2 Å resolution crystal structures [83].

1.2.2 Applications of NMR in the study of protein in solution

After decades of research, NMR spectroscopy is now considered a key technique for structural biology. In fact, the number of protein structures available in the PDB determined using NMR

exceeds 10,000 [84-87]. NMR is not limited to the structure determination of proteins. It is also considered a powerful and versatile spectroscopic tool for the study protein-ligand interactions, dynamics, and protein folding [88-92]. A detailed description of all NMR applications is beyond the scope of this thesis. Therefore, I only highlight two topics pertinent to my work: protein structure determination and protein-ligand interactions.

1.2.2.1 Protein structure determination

Current NMR methods for protein structure determination use molecular dynamics simulations to generate an ensemble of models consistent with the experimental data (e.g, dihedral angles constraints based on coupling constants, dipolar coupling constraints, and intra- and inter-proton distance constraints) and known features of proteins (e.g. bond lengths, and van der Waals radii) [78, 93, 94]. Briefly, strategies to obtain some of the experimental data are discussed below.

Since protein structure determination relies primarily on the ability to obtain interproton distance constraints, near complete resonance assignments are essential. Assignment strategies largely depend on isotopic labeling of the protein [95]. For ^{15}N, ^{13}C labeled proteins, strategies favor the use of 3D experiments (to help reduce spectral overlap) based on through- bond connectives (to help avoid ambiguities in the assignment processes) [95]. For backbone assignments, 3D CBCANH and CBCA(CO)NH experiments correlate the $N(i)/H^N(i)$ chemical shifts of residue (i) with the C_α and C_β chemical shifts of residues (i) + (i-1), and of residue (i-1), respectively [96]. The chemical shifts, together with primary sequence of the protein, can then be used to assign resonances to individual nuclei. For side chain assignments, 3D HBHA(CO)NH, H(CCO)NH, and CC(CO)NH are useful. These experiments correlate the $N(i)/H^N(i)$ chemical shifts of residue (i) to chemical shifts H_α/H_β,$H^{aliphatic}$, and $C^{aliphatic}$ of residue (i-1), respectively [97]. In addition, 3D HCCH TOCSY experiments, which provide chemical shifts for $C^{aliphatic}$ and

$H^{aliphatic}$ on the same residue, can be used to complete or confirm H and C side chain assignments [98].

Once the backbone and side chain resonances have been assigned, interproton constraints can then be obtained using nuclear Overhauser enhancement spectroscopy (NOESY) experiments. Nuclear Overhauser Effect (NOE) result from a distance dependent cross relaxation due to dipole-dipole interactions between NMR observable nuclei. For ^{1}H-^{1}H NOESY experiments, the spectra have two general features: peaks along the diagonal for protons that do not exchange magnetization components and cross-peaks between the coupled protons [99]. These data are qualitatively used to produce NOE interproton distance constraints, which are classified into three distance ranges, 1.8-2.5, 1.8-3.5 and 3-5 Å, corresponding to strong, medium and weak NOEs [76].

Dihedral angle constraints can also be useful. Backbone dihedral angles, can vary from -180˚ to 180 ˚, but due to steric restrictions, they do not adopt all possible values. The allowed ranges of ϕ and ψ angles for different secondary structural elements can be visualized using Ramachandran plots [100]. Since dihedral angles provide important information on the secondary structure of a protein, these values are important for structure generation, and/or to improve structure quality [101]. Several methods to derive torsion angle constraints have been proposed, including three-bond J couplings [76], C_{α}/C_{β},chemical shifts [101], and a database method such as, TALOS+ that predicts torsion angles based on chemical shifts and high-resolution X-ray structures [102]. These strategies all depend on the fact that chemical shifts are highly sensitive to local structure.

Dipolar couplings are distance and angle dependent through-space interactions between two NMR observable nuclei [103]. In solution, the orientation dependent magnetic interactions

average out to zero due to isotropic Brownian motion. In contrast, in weakly aligning media, such as media containing filamentous bacteriophage, anisotropically compressed acrylamide gels, or detergent phases, a non-zero value for the orientation dependent residual dipolar coupling (RDC) can be measured [104, 105]. Since the distance between two nuclei connected by one bond is essentially fixed, for example between ^{15}N and ^{13}C atoms and their attached hydrogens, RDCs provide information about the orientation of the internuclear vectors relative to the protein's magnetic susceptibility axis. RDCs work well in the refinement stage of the structure calculation and increases the precision and accuracy of the structures [103].

Paramagnetic relaxation enhancement (PRE) and pseudocontact shifts (PCS) are useful to obtain long-range distance constraints [106-109]. The PRE effect arises from interactions between the unpaired electron of a paramagnetic spin label and NMR observable nuclei. Due to the large magnetic moment of the paramagnetic center, there is distance dependent (r^{-6}) increase in the relaxation rate of NMR observable nuclei within 15-35 Å. Distance-dependent line broadening effects can then be used to generate long-distance constraints for structure calculations [108]. These experiments work well for proteins with an intrinsic paramagnetic group, or those problems that contain a site specifically conjugated paramagnetic moiety [89].

PCS can also provide long-range information for structure calculations. PCS are observed for paramagnetic ions with an anisotropic magnetic susceptibility tensor χ, such as Fe^{3+}, and lanthanide ions. The large magnetic susceptibility tensor also produces a weak alignment of the protein in the field, resulting in RDCs with long-range information. In contrast to PREs, PCS are governed by a distance dependence of (r^{-3}), to measure longer distances up to ~ 40 Å [89, 109, 110].

1.2.2.2 studies of large proteins

In ^{15}N-^{1}H or ^{13}C-^{1}H correlation experiments the scalar spin-spin coupling splits the signal from each nucleus into two components. As a consequence, in two-dimensional (2D) heteronuclear correlation experiments, a four-line multiplet is observed. Current NMR techniques often collapse the multiplet by a technique called decoupling, resulting in a single, centrally located, averaged signal. For small proteins at high magnetic fields, decoupling of the four-line multiplet results in a simplified spectrum with improved sensitivity. For large proteins, decoupling results in a spectrum with reduced sensitivity, due to the fast transverse relaxation which occurs with a time constant T_2. As a result, assignments strategies for protein above 30 kDa are more difficult [111].

Major sources of transverse relaxation in a heteronuclear spin system include both the dipole-dipole relaxation (DD) and chemical-shift anisotropy (CSA) [112]. The rate of dipole-dipole interactions is field-independent while the chemical shift anisotropy is field-dependent. At high magnetic fields, the CSA can significantly contribute to the transverse relaxation of large proteins [113]. As a result, the components of the four-line multiplet have different linewidths, and decoupling results in a reduced average signal. The linewidths of the four-line multiplet are field-dependent, such that at around 900-1000 MHz (for the ^{15}N-^{1}H moiety) the difference in linewidth of the four-line multiplet is the greatest. Around this magnetic field strength, there is a significant reduction of the transverse relaxation of one of the four-line multiplet, producing a signal of narrow linewidth. In a transverse relaxation-optimized spectroscopy (TROSY) experiment, there is no decoupling and the narrow line of the multiplet is selected.

TROSY experiments can be optimized using partially deuterated proteins in H_2O, where the solvent accessible amide deuterons are exchanged for NMR-observable protons. Since the

relaxation of the of the narrow component of the multiplet is mostly affected by DD interactions with remote hydrogen atoms outside of the ^{15}N-^{1}H moiety [114] the replacement of protons with deuterons reduces proton-proton DD and scalar couplings, resulting in narrower linewidths [115].

Although experiments using partially deuterated proteins have shown promise, deuteration removes most of the side chain protons essential for NOE analysis. To address this challenge, methods have been developed to selectively protonate methyl groups in highly deuterated proteins. With the augmented NOE data set, these methods can be used to study proteins up to 100 kDa [116].

Another approach to study large protein is to use the “divide and conquer” approach, which divides the protein into smaller, more tractable pieces [117]. For these studies, parts a large multi-domain proteins are isotopically labeled and purified separately. Then, if the spectra of these domains are sufficiently similar to the spectra of full-length protein, spectra of the smaller pieces are assigned using traditional ^{15}N, ^{13}C correlation experiments. These assignments can then be transferred to the full-length protein. A similar approach has been utilized to study multi-oligomeric proteins with a high degree of symmetry [118]. For these proteins, the total NMR signals of the full-complex is usually reduced in comparison to a protein of similar size and less symmetry (assuming that each nucleus in the monomeric subunit is usually in the same chemical environment). Mutations that disrupt the oligomeric surface must be found, which is often challenging and may require significant effort. Once the monomeric subunit is isolated, spectra for the monomeric subunit are assigned using conventional labeling and NMR techniques, and assignments are then transferred to the multi-oligomeric protein.

1.2.2.3 Isotopic labeling to study protein-ligand interactions

Solution NMR can be an exceptionally sensitive tool to study protein-ligand interactions. These experiments can be divided into two classes. Protein-observe and ligand-observe NMR. Currently, several informative reviews covering these topics are available [88, 89]. Briefly, in the section below, I include a description of protein-observe experiments for ^{15}N-, ^{13}C- and ^{19}F-labeled proteins.

In protein-observe NMR experiments of protein-ligand interactions, the intensity/chemical shift changes are monitored by titrating ligand into a solution containing isotopically labeled proteins. Chemical shift changes upon ligand binding can result from direct interactions with the ligand, or from long range conformational rearrangements. Therefore, to help reduce ambiguities between local and long range conformational changes, several chemical shift changes are measured and mapped onto the 3D protein structure. Experiments for ^{15}N labeled proteins use 2D ^{1}H-^{15}N HSQC, or TROSY based experiments for larger proteins[89]. Preparation of ^{15}N-labeled protein, expressed using bacterial systems, is now considered relatively straightforward and only requires growth in minimal media supplemented with ^{15}N-labeled ammonium chloride.

Another approach is to use ^{13}C-methyl labeled proteins, which contain ^{13}C probes in the side chain of the protein, such as alanine, valine, isoleucine, leucine and methionine. For ^{13}C-methyl labeled proteins, 2D ^{1}H-^{13}C HMQC experiments are typically used. These experiments have some advantages in comparison to ^{1}H-^{15}N HSQC experiments. Specifically, methyl groups have three protons with a three-fold degeneracy and give rise to stronger signals. They also tend to resonate in a sparsely populated region of the ^{1}H-^{13}C correlation spectrum, reducing spectral overlap [119]. Several suitable labeling schemes for the six canonical methyl-containing amino

acids, including alanine, isoleucine, leucine, methionine, and threonine, have been developed [120]. These labeling schemes often exploit metabolic pathways, metabolic precursors, and the amino acid's propensity for isotopic scrambling at other sites in the protein. For leucine and valine, which share the same metabolic pathway, labeling techniques often result in the incorporation of isotopes into both amino acids. Successful labeling schemes have been carried out by using metabolic precursors common to both amino acids, such as α-ketoisovalerate. Isoleucine, can also be labeled using α-ketobutyrate, one of its metabolic precursor [121-124]. For alanine and methionine, residue specific labeling can be achieved by supplementing minimal expression medium with the appropriate isotopically labeled amino acid. This labeling strategy often results in a considerable amount of isotopic scrambling when labeling alanine residues, since alanine is synthesized from the reversible transamination of pyruvate. Therefore, to reduce isotopic scrambling, the expression media can be supplemented with deuterated amino acids and/or metabolic precursors of amino acid in the relevant metabolic pathways that alanine might contribute to, such as α-ketoisovalerate [117, 125-128]. Unlike alanine, methionine at the end of its metabolic pathway. Therefore, directly supplementing the expression media with ^{13}C-methyl methionine has been found to result in low or undetectable amounts of isotopic scrambling [117].

Although classical ^{1}H, ^{13}C, and ^{15}N spectroscopic approaches have been used extensively to study proteins in solution, ^{19}F NMR is gaining increasing popularity. As discussed previously [129, 130] the 100% naturally abundant ^{19}F atom displays several properties that render it ideal for NMR exploitation: It possesses a spin 1/2 nucleus and a high gyromagnetic ratio that results in excellent sensitivity (83% of ^{1}H). In addition, the shielding of the ^{19}F nucleus is dominated by a large paramagnetic term and, as a result, fluorine chemical shifts are exquisitely sensitive to changes in local environment (the chemical shift range is ~100-fold larger than that of ^{1}H).

Another great advantage of using ^{19}F as an NMR probe is its absence from virtually all naturally occurring biomolecules, small and large. For this reason, studies of fluorinated biopolymers can be carried out in any routinely used buffer system or environment without suffering from interference by background signals. Thus, no special precautions are needed to remove buffer and additive signal intensity from the spectra. The van der Waals radius of the ^{19}F atom (1.47 Å) lies between those of hydrogen (1.2 Å) and oxygen (1.52 Å), and strategic substitution of ^{19}F atoms for hydrogens, hydroxyl groups, or carbonyl oxygens in biological molecules is considered weakly perturbing and often has little effect on a protein's biological activity [129-132].

For 40 years now, the ^{19}F atom has been exploited in biological NMR [133, 134], and several reviews have covered the field over the last four decades [129-131, 135, 136]. Indeed, the ^{19}F atom has been used as a molecular probe to gain insight into protein and peptide structure [129, 130, 137], protein-ligand interactions [129-132, 138, 139], protein unfolding [133, 134, 140, 141], protein aggregation [129-131, 135, 136, 142, 143], and protein dynamics [144], clearly demonstrating the power and versatility of the fluorine probe for NMR.

Several methods to prepare ^{19}F-modified proteins have been described [129, 131, 145-148]. These methods fall into three main categories: (i) post-translational covalent attachment of ^{19}F-containing moieties to the protein, (ii) biosynthetic amino acid type-specific incorporation of ^{19}F-modified amino acids, and (iii) site-specific incorporation of ^{19}F-modified amino acids using recombinantly expressed orthogonal amber tRNA/tRNA synthetase pairs.

In brief, post-translational covalent modification introduces ^{19}F atoms into the protein of interest by conjugating a ^{19}F-containing moiety to a reactive group, such as an -SH group on a solvent accessible cysteine [137, 138]. One advantage of this technique is the ability to

incorporate the label into proteins for which biosynthetic labeling is cost-prohibitive, such as proteins expressed in mammalian cells. For residue-specific incorporation of ^{19}F-modified amino acids into proteins, expression is carried out in defined growth media supplemented with the ^{19}F-modified amino acid. This method relies on the endogenous aminoacyl-tRNA synthetases to charge the ^{19}F-modified amino acid onto their cognate tRNAs. As a result, all codons recognized by the amino acid-specific tRNA will carry the ^{19}F-modified amino acid, and global incorporation of the ^{19}F-modified amino acid into all proteins will occur. To maximize the efficiency of ^{19}F-labeled amino acid incorporation, auxotrophic bacterial strains that cannot synthesize the amino acid that is to be replaced are sometimes used, but frequently are not necessary [149, 150]. Still, for some amino acids, >90% labeling can be achieved if conditions are worked out carefully. Potentially the most powerful and versatile method to introduce ^{19}F atoms into proteins is site-specific incorporation of ^{19}F-modified amino acids using a recombinantly introduced orthogonal amber tRNA/tRNA synthetase pair. This method is based on an extension of the genetic code to beyond the natural 20 amino acids, first described by Noren and colleagues in 1989 [151]. It uses nonsense stop codons and nonsense suppressor tRNAs to overturn termination of protein biosynthesis and requires engineered aminoacyl-tRNA synthetases that specifically acylate the suppressor-tRNAs with the non-natural ^{19}F-modified amino acid *in vivo*, without interfering with other tRNA/synthetase pairs. Application of this approach entails the introduction of the amber nonsense codon (TAG) at any desired location in the protein coding sequence, such that it replaces the natural amino acid codon in that location, in conjunction with introduction of a tailored orthogonal amber tRNA/tRNA synthetase pair. More specifically, a vector containing the protein amber-mutant gene is co-introduced into an *E.coli* host strain along with a vector encoding an *in vitro* evolved orthogonal amber tRNA/tRNA

synthetase pair that recognizes the ^{19}F-modified amino acid. Expression is then carried out in growth media supplemented with the ^{19}F-modified amino acid. One drawback of this methodology is the fact that the orthogonal amber tRNA/tRNA synthetase pair does not incorporate the ^{19}F-modified amino acid with 100% efficiency at the amber codon and that translation termination occurs to varying degrees. As a result, two protein products are invariably generated, an undesired unlabeled truncated protein, and a 100% labeled full-length protein. These two proteins can be separated during purification if a C-terminal affinity tag is used. All three methods are schematically illustrated in **Figure 1.4**. The fluorine labeling section was adapted from [153].

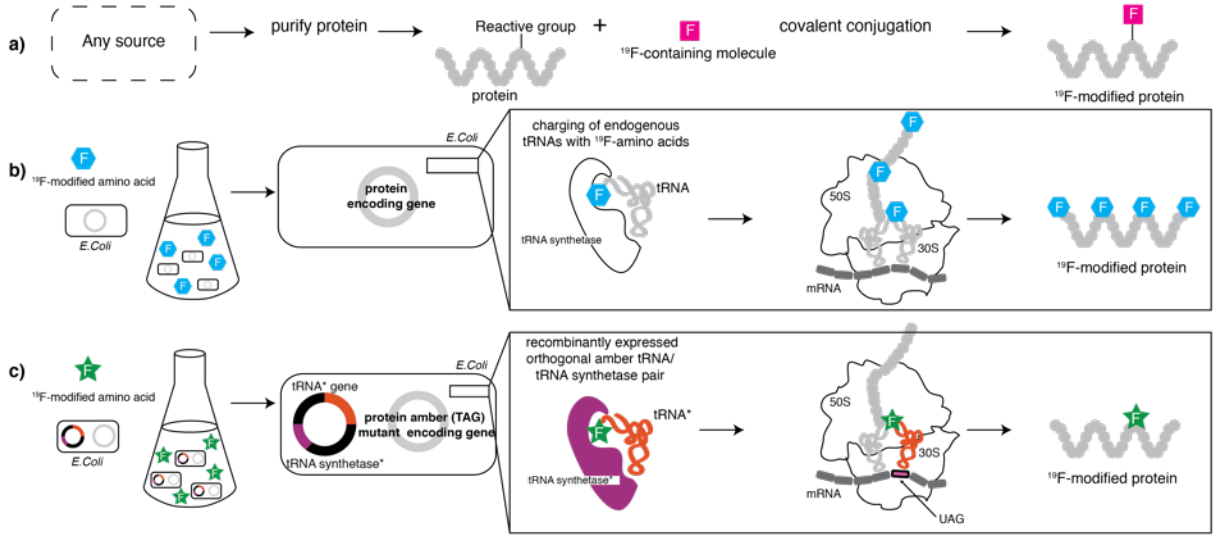

Figure 1.4 The three main methods to prepare ^{19}F-modified proteins are shown.

a) Posttranslational covalent conjugation of ^{19}F-containing moieties to the protein. b) Biosynthetic amino acid type-specific incorporation of ^{19}F-modified amino acids. c) Site-specific incorporation of ^{19}F-modified amino acids using recombinantly expressed orthogonal amber tRNA/tRNA synthetase pairs.

2.0 SUMMARY OF PROJECTS

Solution NMR is powerful for *de novo* structure determination and to obtain site-specific informational on large proteins. Given RT's size, conventional NMR experiments that use protein uniformly labeled with ^{13}C and ^{15}N are challenging. Therefore, we use several different strategies to study RT in solution. In particular, we used the "divide and conquer" approach to study the p66 immature precursor of RT (**Chapter 5.0**). We also use the "divide and conquer" approach combined with traditional NMR methods for structure determination of the isolated thumb subdomain of RT (**Chapter 6.0**). Furthermore, we selectively introduced and trifluoromethyl phenylalanine (tfmF) to study RT-inhibitor interactions (**Chapter 7.0**). Together, our studies provide important novel information regarding RT maturation, the structure of the HIV-1 thumb subdomain, and RT-inhibitor interactions.

3.0 PROTEIN EXPRESSION AND PURIFICATION.

All the studies in this thesis use similar protocols to produce and purify RT proteins. Therefore, in an effort to reduce redundancy, all these procedures have been grouped into one section. This chapter includes the construction of the protein coding genes **section 3.1**, protocols used for producing unlabeled and isotopically labeled proteins **section 3.2**, and purification protocols for RT proteins, subunits and (sub)domains **section 3.3**.

3.1 CLONING

The coding sequence for the RT p66 subunit was amplified from the p6HRT-PROT vector, kindly provided by Dr. Sluis-Cremer, using 5'-acc gca cat atg ccc att agc cct att gag act gta-3' and 5'-gca gat ctc gag tag tat ttt cct gat tcc agc act gac-3' as forward and backward primers, respectively. The amplified product was inserted into the pET21a(+) vector (Invitrogen, Carlsbad, CA), which encodes a six histidine tag at the C-terminus of the protein construct. After initial expression and purification trials, a codon-optimized C280S/C38V double cysteine version was created for increased protein expression in *E. coli* (DNA 2.0 gene synthesis, Menlo Park, CA). The coding sequence for the p51 subunit with an N-terminal Strep-tag was created by amplification of the appropriate region (coding for residues 1-440) of the p66 RT codon optimized sequence, using 5'-cc gca tcc atg gat tgg agt cac ccg cag ttc gag aaa cca atc agc cca atc

gaa acg gtc cc -3’ and 5’-ccg cat ctc gag tta gaa cgt ttc cgc gcc aac aat cgg ttc ttt ctc c-3’ as forward and reverse primers, respectively, and the amplified product was then inserted into a pET28a+ vector (Invitrogen). Constructs coding for residues 1-216 (Finger-Palm domain), 237-318 (Thumb domain) and 427-556 (RNH domain) were amplified from the p66 sequence using 5’-gca gct cat atg cca atc agc cca atc gaa acg gtc cc-3’ and 5’-gca gat ctc gag ggt cgt cag acc cca acg cag cag atg c-3’ as forward and backward primers, respectively, for the finger-palm domain, 5’-cgt acg cat atg gat aaa tgg aca gta cag cct ata gtg ctg cc-3’ and 5’-cgt acg ctc gag ata cac tcc atg tac tgg ttc ttt tag aat ctc-3’ as forward and backward primers, respectively, for the thumb domain, and 5’-gca gat cat atg tat caa ctg gag aaa gaa ccg att gtt ggc-3’ and 5’-gca gat ctc gag cag gat ttt gcg aat acc tgc gct cac c-3’ as the forward and backward primers, respectively, for the RNH domain. See **figure 3**, for visual a representation of the amplified p66 coding regions used to produce RT subunits and (sub)domains.

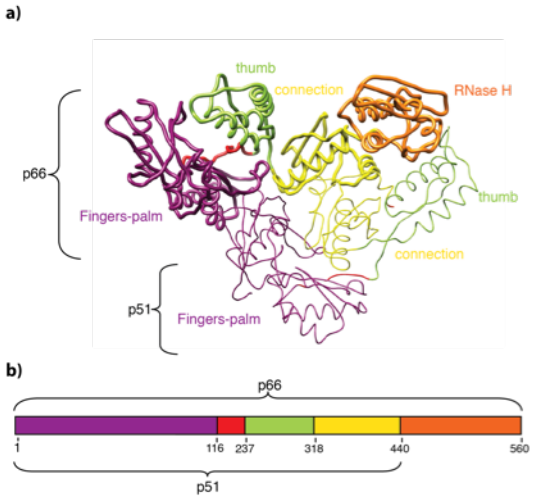

Figure 3.1 General description of RT structure and schematic representations of the p66 coding regions amplified to produce RT subunits and (sub)domains.

(a) Tube representation of HIV-1 RT. b) Schematic representation of the p66 gene, labeled to indicate the coding regions for RT proteins. The fingers-palm, thumb, connection, and RNH (sub) domains in the p66 and p51 subunits colored in magenta, green, yellow and orange, respectively

For singly 4-trifluoromethyl phenylalanine (tfmF)-labeled proteins, the vectors encoding p66-127tfmF, p66-146tfmF, and p66-181tfmF proteins, were generated by replacing codons for tyrosines at positions 127, 146, or 181 with amber codons, using the appropriate TAG oligonucleotides as primers, and the p66 encoding vector as a template. The vectors encoding mutant p66-181tfmF, p66-146tfmF and p51 proteins (p66-181tfmF-K103N, p66-181tfmF-V108I, p66-146tfmF-K103N, p66-146tfmF-V108I, and the p51-V108I, p51-E138K, p51-K103N) were generated using the appropriate oligonucleotides as primers, and the p66-181tfmF, p66-146tfmF, and p51 encoding vectors as templates, respectively. All the p66, thumb, RNH, and finger-palm constructs contain a hexa-histidine tag at the C-terminus, and all the p51

constructs contain a Strep-tag at the N-terminus. The DNA sequences of all constructs were verified by DNA sequencing (Genewiz, South Plainfield, NJ).

3.2 PROTEIN PRODUCTION

Unlabeled proteins (p51, p51-V108I, p51-K103N, p51-E138K) were produced in *E. coli* BL21 (DE3) gold cells (Agilent Technologies, Santa Clara, CA**)**, using 0.5 mM IPTG for induction over 16 hours at 27ºC or by growing cells for 24 h 27ºC using auto-induction medium [152, 153].

Uniform ^{15}N-labeling of the isolated Finger-Palm, Thumb, and RNH domains was achieved using modified minimal medium containing $^{15}NH_4Cl$ as the nitrogen source. Uniform ^{15}N- and ^{13}C-labeling of the Thumb and RNH domains, was carried out, using $^{15}NH_4Cl$ and $^{13}C_6$-glucose as sole nitrogen and carbon sources, respectively. Uniform ^{15}N- and ^{2}H- labeling of the p66 and the p51 protein was achieved in modified minimum medium containing $^{2}H_2O$ and $^{15}NH_4Cl$.

The tfmF labeled p66 proteins (p66-127tfmF, p66-146tfmF, p66-181tfmF, p66-181tfmF-V108I, and p66-181tfmF-K103N) were produced using a protocol developed by the Mehl laboratory [147, 153]. Briefly, competent *E.coli* BL21 ai cells (Invitrogen) were co-transformed with the vector encoding the TAG containing constructs (above), and the pDule2 RS vector encoding the orthogonal amber tRNA/tRNA synthetase pair. Several transformants were screened for expression. All tfmF containing p66 proteins were produced at 27°C by growing cells for 24 h using auto-induction medium, containing tfmF at a final concentration of 1 mM.

3.3 PROTEIN PURIFICATION

Cells were harvested by centrifugation at 5887 *g*, re-suspended in lysis buffer containing 25 mM sodium phosphate (pH 7.5), 25 mM imidazole and 500 mM NaCl, and lysed using a microfluidizer. Cell debris was removed by centrifugation at 34530 *g*. Purification of RT proteins generally involve the use of a combination of four columns: 5 mL HisTrap column, StrepTrap column, 5 mL HiTrap SP column and HiLoad 26/60 Superdex 200 column (GE Healthcare). For purification steps using the HisTrap column, the column was equilibrated with lysis buffer and eluted with a linear gradient of 0.025-0.5 M imidazole (except for p51 proteins). For purification steps using the StrepTrap column, the column equilibrated in 25 mM sodium phosphate, 6 mM KCl, 280 mM NaCl (pH 7.5) and proteins were eluted using the same buffer listed above supplemented with D-desthiobiotin to a final concentration of 3 mM (Sigma Aldrich, St. Louis, MO). For purification steps using the HiTrap SP column, the column was equilibrated with 25 mM sodium phosphate (pH 6.5), and eluted using a linear gradient of 0-0.5 M NaCl. For purification steps using the HiLoad 26/60 Superdex 200 gel filtration column, phosphate-buffered saline was used.

The thumb, RNH and finger-palm (sub)domains and the p66 proteins, which contain a C-terminal His-tag, were purified over a 5 mL HisTrap column, followed by 5 mL HiTrap SP column.

For the p51 proteins, which contained an N-terminal Strep-tag, the supernatant after cell lysis and centrifugation was loaded onto a 5 mL HisTrap column. The flow through was collected and further purified over a 5 mL StrepTrap column, followed by a 5 mL HiTrap SP column.

For RT proteins, equivalent amounts of cell pellets, containing the expressed p66 and p51 proteins, were mixed, and lysed using a microfluidizer. After centrifugation, the supernant was applied onto a 5 mL HisTrap column. The elution with a linear gradient of 0.025-0.5 M imidazole was essential to separate fractions containing p66/p51 heterodimeric RT proteins from p51 monomer, and p66 proteins. RT-containing fractions further purified over 5 mL StrepTrap column (GE healthcare), a step essential for the removal of p66 homodimer.

The final purification step for all protein samples included the use a HiLoad 26/60 Superdex 200 column (GE Healthcare). Protein fractions were pooled and concentrated in an Amicon Ultra concentrator (EMD Millipore, Billerica, MA) to ~10 µM. Glycerol (50% v/v glycerol) was added and samples were stored at -20°C for future use. A list of proteins used in the following chapters is provided in **Table 3.1**.

Table 3.1 RT proteins

Proteins	Subunits	Described in Chapter
^{15}N thumb subdomain	thumb subdomain	4
^{15}N finger-palm subdomain	finger-palm subdomain	4
^{15}N RNH	RNH domain	4
^{15}N p51	p51	4
^{2}H,^{15}N p66	p66	4
^{13}C, ^{15}N thumb subdomain	thumb subdomain	4,5
RT127tmfF	p51/p66-127tfmF	6
RT146tfmF	p51/p66-146tfmF	6
RT181tfmF	p51/p66-181tfmF	6
RT181tfmF-V108I	p51-V108I/p66-181tfmF-V108I	6
RT181tfmF-K103N	p51-K103N/p66-181tfmF-K103N	6
RT181tfmF-E138K(p51)	p51-E138K/p66-181tfmF	6
RT146tfmF-V108I	p51-V108I/p66-146tfmF-V108I	6
RT146tfmF-K103N	p51-K103N/p66-146tfmF-K103N	6

4.0 THE P66 PRECURSOR OF HIV-1 RT

In contrast to the wealth of structural data available for the mature p66/p51 heterodimeric human immunodeficiency virus type 1 reverse transcriptase (RT), the structure of the homodimeric p66 precursor remains unknown. In all X-ray structures of mature RT, free or complexed, the processing site in the p66 subunit, for generating the p51 subunit, is sequestered into a β-strand within the folded ribonuclease H (RNH) domain and is not readily accessible to proteolysis, rendering it difficult to propose a simple and straightforward mechanism of the maturation step. Here, we investigated, by solution NMR, the conformation of the RT p66 homodimer. Our data demonstrate that the RNH and Thumb domains in the p66 homodimer are folded and possess conformations very similar to those in mature RT. This finding suggests that maturation models which invoke a complete or predominantly unfolded RNH domain are unlikely. The present study lays the foundation for further in-depth mechanistic investigations at the atomic level. The results presented in this chapter have been adapted from: Sharaf NG, Poliner E, Slack RL, Christen, MT, Byeon, I-JL, Parniak, MA, Gronenborn, AM, Ishima, R (2014) The p66 immature precursor of HIV-1 reverse transcriptase. Proteins 82:2343–2352.

4.1 INTRODUCTION

RT plays a central role in the replication of all retroviruses and related retrotransposons. In the HIV-1 life cycle, RT is expressed as part of the Gag-Pol polypeptide, which is processed by retroviral protease into several proteins [154-158]. The mature RT enzyme is a heterodimer, composed of two subunits, p66 and p51 (**Figure 4.1a**) [159-162]. The RT p51 subunit is generated by removal of the C-terminal RNH domain from p66 by PR [163-167]. *In vivo* processing of the Gag-Pol polyprotein is complex, and the detailed mechanism of RT maturation into the heterodimer is still unclear. Based on data obtained from model systems, cleavage at the p51-RNH processing site is assumed to occur in a p66 homodimer [36, 164, 168]. However, in all known RT X-ray structures, as well as those of the isolated RNH domain, the p51-RNH cleavage site is located within the folded RNH domain, sequestered into the center of a β-sheet, and thus seemingly inaccessible to the protease (**Figure 4.1b**) [36, 39, 169-171]. No significant motions were observed at the p51-RNH processing site in the isolated RNH domain [172, 173], which may have suggested partial accessibility of the site. In addition, the lack of structural information on the "immature" p66 RT precursor renders any mechanistic explanation(s) tentative. We, therefore, embarked on studies aimed at providing the foundation for structurally elucidating RT processing.

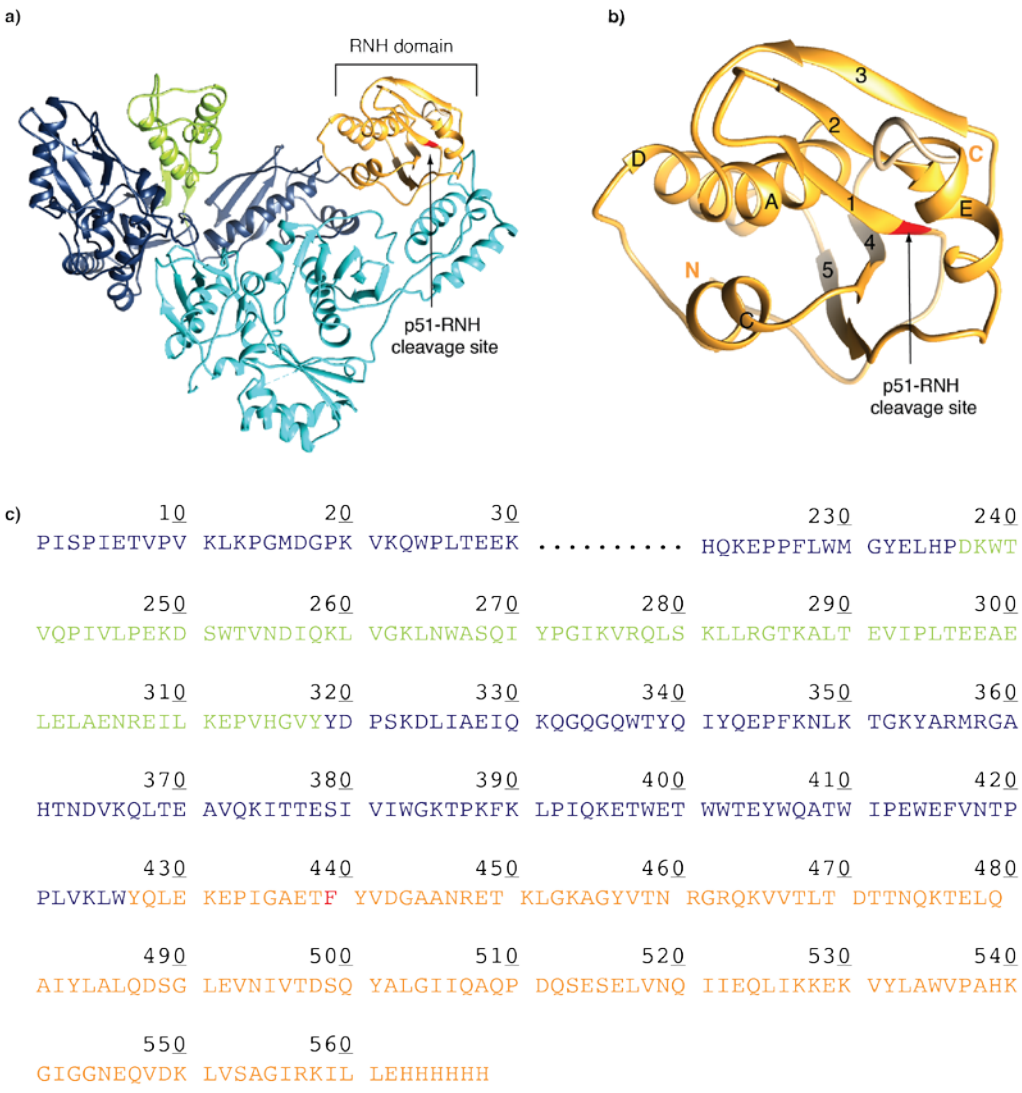

Figure 4.1 Ribbon representation of the structures RT and RNH domain

a) p66/p51 RT heterodimer and b) the RNH domain, indicating the p51-RNH processing site (arrow), and c) amino acid sequence of p66. In a)-c), the Thumb and RNH domains in the p66 subunit are show

To evaluate protein conformation in solution, Nuclear Magnetic Resonance (NMR) spectroscopy provides powerful approaches [85], since it permits the investigation of conformational equilibria and protein dynamics at the amino acid residue level [174-176].

However, NMR studies of HIV RT are challenging, given the protein's large molecular mass (117 kDa); to date, NMR of RT has been mostly limited to observing methyl groups of side chains, such as in methionine or isoleucine [177-179]. Although methyl resonances are valuable probes for obtaining general qualitative information about a protein's conformation in solution [124, 180], they report only on a limited number of positions and, therefore, cannot inform on the secondary and tertiary structural details that are mirrored by a protein's backbone chemical shifts [181].

Here, we present an investigation of the p66 homodimeric RT. The p66 dimer possesses enzymatic activity [182, 183] and is widely considered to function as the RT precursor [36, 164, 168, 184]. We took advantage of the extreme sensitivity of backbone amide resonance frequencies to assess conformational similarities between different protein constructs. In the ^{1}H-^{15}N heteronuclear single-quantum coherence (HSQC) spectrum of the p66 homodimer, over 240 resonances were observed. Comparison of the p66 spectrum with the spectra of the isolated domains revealed that greater than 60 % of the isolated Thumb domain and more than 40 % of the isolated RNH domain resonances, respectively, are in very similar positions. In contrast, only 18 % of the Finger-Palm domain resonances match those of the p66 homodimer. This establishes that both the Thumb and RNH domains are stably folded in the immature p66 homodimeric RT and exhibit essentially the same structures as in the isolated domains. With these findings in mind, the question arises how HIV-1 protease gains access to the p51-RNH processing site in the p66 homodimer. Our data suggests maturation models that invoke a complete unfolded or predominantly unfolded RNH domain [169, 170] are unlikely, and implies that p51-RNH processing may involve selection of a minor conformation or a protease-binding induced structure, which is cleaved during maturation.

4.2 EXPERIMENTAL PROCEDURES

4.2.1 NMR experiments

Samples were buffer-exchanged into NMR buffer (25 mM sodium phosphate, pH 6.8, 100 mM NaCl) in an Amicon Ultra concentrator (EMD Millipore). In the final exchange step, the sample was concentrated to 350 µL and supplemented with 10% (v/v) D_2O. Final protein concentrations for NMR experiments were ~200 µM (in monomer). All NMR spectra were recorded at 303 K on Bruker 600 and 900 MHz AVANCE spectrometers, equipped with 5-mm triple-resonance, *z*-axis gradient cryoprobes. For p66 and p51 proteins, a TROSY version the ^{1}H-^{15}N HSQC NMR experiments was used [114, 185]. For backbone assignments of the thumb and RNH domains 3D CBCAONH, HNCACB, and ^{1}H-^{15}N NOESY-HSQC [96] spectra were recorded. All data were processed with NMRPipe and analyzed with CCPN analysis [186, 187]. Chemical shifts of amide resonances, as defined by **Equation 1**, were deemed identical if the picked peaks in two spectra resided within ± 0.03 ppm.,

$$\Delta\delta^{i}_{obs} = \sqrt{\left(\delta^{i}_{H} - \delta^{free}_{H}\right)^{2} + \left[\left(\delta^{i}_{N} - \delta^{free}_{N}\right) \times \frac{\gamma_N}{\gamma_H}\right]^{2}} \qquad (1)$$

Here, δ indicates chemical shift in ppm units, and γ_N and γ_H indicate gyromagnetic ratios of N and H resonances, respectively. Secondary structure elements were delineated using CSI and TALOS+ [102, 188]. Chemical shifts for the Thumb domain spectrum were estimated using the coordinates from the crystal structure for residues 237 to 318 of p66 in RT p66/p51 (PDB ID: 1DLO), with the program Sparta [189].

4.2.2 Multi-angle Light Scattering

Size-exclusion chromatography/multi-angle light scattering (SEC-MALS) data were obtained at room temperature using an analytical Superdex 200 (10 × 300 mm, GE Healthcare) column with in-line multi-angle light scattering, refractive index (Wyatt Technology, Inc., Santa Barbara, CA) and UV (Agilent Technologies) detectors. 100 µL of a protein solution (10.9 - 59.9 µM for p66 and 19.3 - 58.6 µM for p51) was loaded onto the column, pre-equilibrated and eluted with NMR buffer at a flow rate of 0.5 mL/min. The molecular masses of the eluted protein species were determined using the ASTRA V.5.3.4 program (Wyatt Technologies). The dimer association constant for p66/p66 was extracted from peak intensities at the monomer and dimer molecular masses.

4.3 RESULTS

4.3.1 Amide backbone resonances of homodimeric p66

To obtain insight into the conformation of the immature p66 homodimeric RT in solution, ^{1}H-^{15}N TROSY HSQC spectra of perdeuterated p66 were recorded (**Figure 4.2a**). Excluding resonances that likely arise from amino acid side chains, at least 247 resonances were resolved. Although this number is fewer than half of all possible amide resonances (~600), it is still remarkable that such a large number is observed, given the large molecular mass of the p66 homodimer and concomitant line broadening in TROSY HSQC spectra [180, 190]. Note that the p66 sample is predominantly a homodimer (>80 %) at the concentration used (~200 µM), based

on multi-angle light scattering (MALS) analysis (**Figure 4.3a**) and a previously reported dissociation constant, K_D, of 4 μM [163, 165].

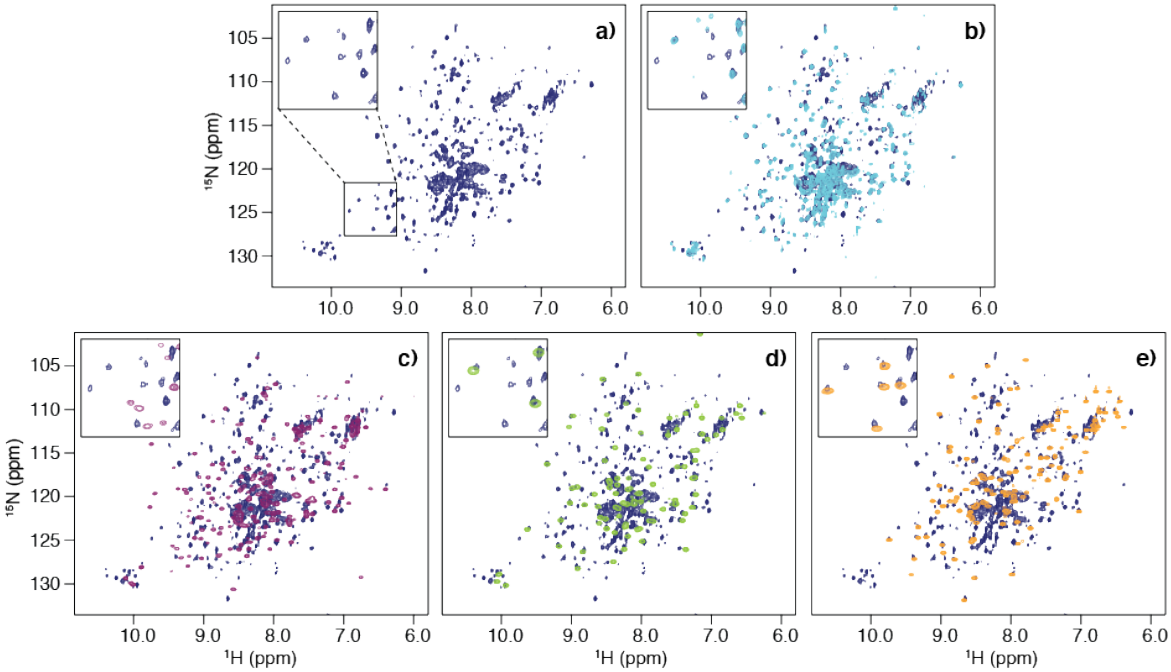

Figure 4.2 ^{1}H-^{15}N HSQC spectra

a) p66 (dark blue), b) p51 (cyan), c) palm-finger (purple), d) thumb (green), and e) RNH (orange). In b)-e), individual isolated domain spectra are superimposed on the p66 spectrum (dark blue). The insets depict expanded regions of the spectra as indicated in panel a).

For comparison, a ^{1}H-^{15}N TROSY HSQC spectrum was also recorded for the p51 sample (**Figure 4.2b**), which contained less than 60 % dimer, based on MALS analysis (**Figure 4.3**) and a reported K_D of 230 μM [163, 165]. Given that both our p66 and p51 NMR samples contain monomeric and dimeric protein, but give rise to only one set of resonances, the monomer and dimer forms are likely in fast exchange on the chemical shift scale. For p51, with a K_D >>1 μM and an estimated on-rate >10^6 s^{-1}/M, [191, 192] this is not surprising. For p66, with a K_D ~4 μM,

the observed chemical shift positions are most likely those of the major dimer species (>80 %), irrespective of the exchange régime.

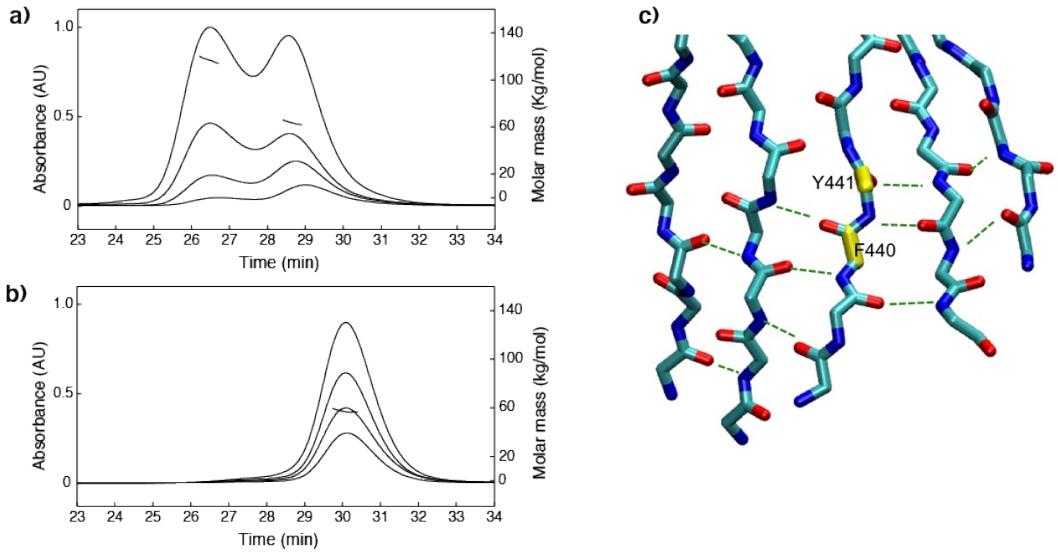

Figure 4.3 Multi-angle light scattering (MALS) elution profiles and a portion of the RNH backbone structure with a subset of hydrogen bonds highlighted (dashed lines).

a) the p66 and b) the p51 protein samples. Experiments were repeated at different protein concentrations from 10.9 μM to 59.9 μM and 19.3 μM to 58.6 μM, for p66 and p51 samples, respectively. Assuming that the samples are diluted approximately 10-fold on the gel-filtration column, homodimer dissociation constants for p66 and p51, calculated from the observed MALS results, are 4.3 ± 0.7 μM and 318 ± 116 μM, respectively, which are in excellent agreement with the published homodimer dissociation constants of ~4 μM and ~230 μM for p66 and p51, respectively. c) We expect that at least these hydrogen bonds exist in the p66 homodimer because there are amide cross peaks in the ^{1}H-^{15}N TROSY spectrum of the p66 homodimer at positions similar to those of the isolated RNH fragment (<27 Hz). NMR cross peaks that are located in overlapped regions could not be evaluated here. The NMR structure of the isolated RNH fragment (PDB = 1O1W [193]) was used to generate the graphics.

Despite the different proportion of dimerization in the p66 and p51 samples, a significant number of amide resonances exhibit the same resonance frequencies in the p51 and p66 NMR spectra; this is easily seen upon superposition of the spectra (**Figure 4.2b** and **Table 4.1**). Since NMR chemical shifts are extremely sensitive to local environments, those associated with residues that experience a different environment in two states, for example when located at domain interfaces in multimers, are expected not to be the same. Thus, the observed high similarity of the spectra, irrespective of the difference in dimer population, suggests that the observed signals represent residues that reside in identical local structures of p51 and p66. Interestingly, the fact that we observe a large number of resonances in the ^{1}H-^{15}N TROSY HSQC spectrum for the p66 homodimer, despite its large molecular mass, suggests that maybe some of the smaller domains exhibit a certain degree of independent motion, faster than the overall rotational molecular diffusion.

4.3.2 The RNH and Thumb domains are independently folded domains in the p66 homodimer

The RNH domain of HIV-RT can exist as a stably folded, individual domain with a structure similar to that in the intact protein [169, 173, 193]. It is not known, however, whether other RT domains can also exist as stable folded sub-structures. We, therefore, prepared the Finger-Palm domain (residues1 to 216), the Thumb domain (residues 237 to 318) and the RNH domain (residues 427 to 556, see **Figure 4.1c**). The ^{1}H-^{15}N HSQC spectra for all three isolated domains exhibit well-dispersed resonances, indicative of stably folded structures (**Figure 4.2c-e**). To qualitatively assess whether the domain structures within p66 and in isolation are identical, the

isolated domain spectra were individually superimposed onto the spectrum of p66 (**Figure 4.2c-e**). Amide resonances that were considered identical in frequencies (0.03 ppm in the combined chemical shift in **Equation 1**) were counted: of the three domains, the isolated Thumb and RNH domains exhibited large degrees of identity with the intact p66 protein (>40%), while for the finger-palm domain it was low (18%, **Table 4.1**). When comparing the spectra of the individual domains with their counterparts in p51, a high degree of resonance frequency identity was observed for the thumb domain (55 %). We also evaluated whether, by chance, any spectral similarity exists between the structurally diverse RNH and thumb domains; gratifyingly no overlap was noted, as can be appreciated from the superposition of their spectra onto the p66 spectrum (compare the orange and green resonances in **Figure 4.4a**). Likewise, p51 and RNH domain resonances do not coincide (18 % identity, compare the cyan and orange resonances), whereas significant overlap occurs for p51 and p66 (compare the cyan and dark blue resonances in **Figure 4.4b**). Overall, the superposition up to 20 % is explained by signal overlap at the crowded regions in the spectra while those above 20 %, such as comparison of the isolated RNH domain signals with those in p66 likely indicates structural similarity between the two.

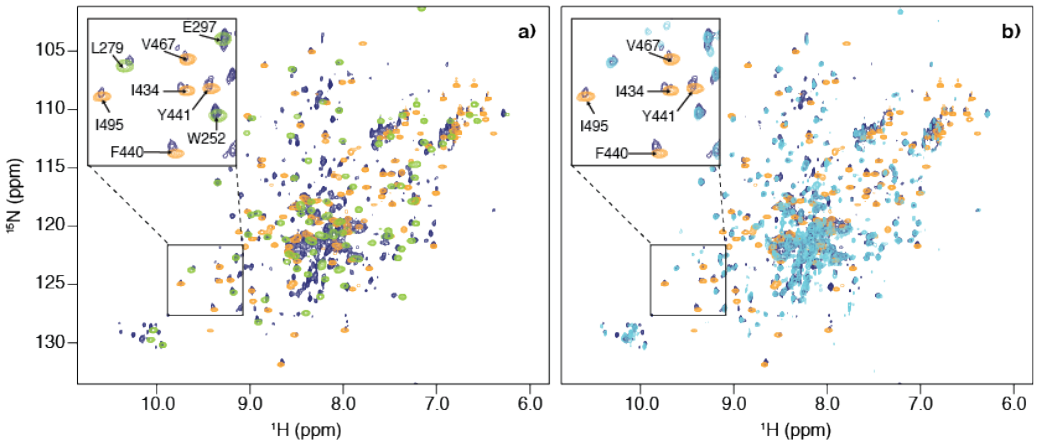

Figure 4.4 Superposition of ^{1}H-^{15}N HSQC

a) of the thumb domain (green) and of the RNH domain (orange) onto the p66 spectrum (dark blue) and b) of the p51 domain (cyan) and the RNH domain (orange) onto the p66 spectrum (dark blue). Several resonances are labeled with amino acid names and numbers in the insets.

Since no structure is available for the isolated Thumb domain and it was not clear a priori whether the isolated domain has the same conformation as the Thumb domain in the p66 homodimer, we used NMR to investigate this question. Backbone assignments were obtained for a sample of the isolated domain and qualitatively assessed using CSI and TALOS+ [102, 181, 188]. The derived secondary structure elements (α-helices) are consistent with an overall domain structure similar to the one observed in the p66/p51 RT crystal. Likewise, Cα chemical shifts that were predicted using Sparta, [189] based on the coordinates of the Thumb domain in the RT heterodimer crystal structure (PDB = 1DLO [38]), match the experimental ones of the isolated Thumb domain (**Chapter 5**). The similarity in secondary structure elements, and probably the tertiary structure, for the Thumb domain alone and in the p66 homodimer is further inferred from the resonance overlap noted above (59%, **Table 4.1**).

Table 4.1 Percentage of amide resonances in the spectra of individual isolated domains that reside at identical frequencies in the p66 or p51 Spectrum

	p51	RNH	thumb	finger-palm
Total number of counted resonances in each spectrum	219	119	83	179
Percentage of signals in the individual domain spectra that coincide with those in p66	62 %	40 %	59 %	18 %
Percentage of signals in the individual domain spectra that coincide with those in p51	-	18 %	55 %	18 %

Resonances were assumed to be identical if they resonated within 0.03 ppm (see "Experimental Methods").

The combined results imply that the Thumb and RNH domains within the p66 homodimer and the Thumb domain in the p51 exhibit the same overall structures as the isolated domains. This, however, is not the case for the Finger-Palm region (**Table 4.1**). For this domain, notable chemical shift differences were observed in the p66 homodimer, possibly due to local structural changes or altered domain-domain or subunit-subunit interactions.

4.3.3 Conservation of the p51-RNH processing site conformation in p66

The conformation of the p51-RNH processing site (F440↓Y441) in the p66 homodimer was assessed based on the comparison of the amide resonance frequencies in the ^{1}H-^{15}N HSQC spectrum of the p66 homodimer and those in the isolated RNH domain [193, 194]. Backbone amide resonances from residues in the β-1 strand, which contains the processing site in the isolated RNH domain, were essentially identical (amino acids 438 to 447) and several of these

(I434, F440, and Y441) are labeled in the **Figure 4.4b** inset. The fact that no chemical shift differences were noted for these processing site residues suggests that they are similarly located in a β-strand within the p66 homodimer **Figure 4.4**. Since NMR resonance frequencies are highly sensitive to local electronic and conformational influences, the close match between amide resonances for the p51-RNH processing site in the p66 homodimer and the isolated RNH domain indicates that the site is similarly structured in both. We also did not observe a significant number of random coil resonances in the p66 spectrum, making it very unlikely that one or both of the RNH domains is unfolded or disordered in the RT p66 homodimer. Since peak volumes of these RNH signals are similar or slightly larger than those of the observed indole NH signals in the p51 domain in the p66 spectrum (**Figure 4.4**), a probability that another RNH conformation, than that identified in our inspection, exists in the p66 homodimer is low. Even with possible exchange between an unfolded and folded RNH domain in p66, any unstructured fraction would be <5% of total, given the close match between resonance frequencies for the bona-fide folded, isolated RNH domain and the RNH domain in p66. Given all of the above, it is strongly suggested that the processing mechanism has to involve the selection of a very minor conformer in the overall conformational ensemble by the protease or a local conformational change upon protease binding.

4.4 DISCUSSION

Numerous p66/p51 RT structures are available, with and without substrate, inhibitors and other ligands, allowing for a detailed elucidation of the conformational transitions that are possible in the RT heterodimer upon ligand interaction. In all of these structures, differences in domain

orientations between the two subunits are observed [36, 39, 170, 195]. In contrast, structural information for the p66 homodimer, which is assumed to be the precursor of mature RT [36, 164, 168, 184], is not available. Therefore, any conformational changes that accompany processing of the homodimer into the heterodimer remain ambiguous.

Here, we investigated the solution conformation of the p66 homodimer, aimed at providing the basis for elucidating RT heterodimer formation. We showed that the ^{1}H-^{15}N TROSY HSQC spectrum of the p66 homodimer contains a very large number of backbone amide resonances, a rather surprising finding for a protein with a molecular mass of 132 kDa. This suggests that some of the smaller domains exhibit a certain degree of independent motion, faster than the overall rotational molecular diffusion. Our study also establishes that the overall major structures of the Thumb and RNH domains within the p66 homodimer are very similar to those of the isolated domains. In addition, the high similarity in chemical shifts for the p51-RNH processing site in the isolated RNH domain and the p66 homodimeric RT indicates that the cleavage site in the p66 homodimer exhibits essentially the same conformation as the isolated RNH domain, seemingly buried in the β-structure of the domain.

The mechanistic details involved in the formation of the mature RT p66/p51 heterodimer are not well understood, but several models have been proposed. One prevailing model assumes that cleavage of the immature p66 homodimer occurs in an unstructured RNH domain (**Figure 4.1c**) [169, 170], with subsequent folding of the remaining p66-RNH domain into the structure of the mature heterodimeric RT. This hypothesis is based primarily on the assumption that p51-RNH processing would be very inefficient if the processing site were located within a structured domain [196]. Another proposal assumes that the p66 homodimer exists in an asymmetric conformation [166, 168]. Indeed, the p51 domain conformation in the p66 subunit differs from

that in the p51 subunit of the mature RT heterodimer, and this could be true in the p66 precursor as well. A third proposal suggests cleavage of p66 monomers, followed by folding of both p66 and p51 monomers, and formation of a mature p66/p51 heterodimeric RT (a concerted model) [163, 164]. While our NMR results do not allow us to propose an exclusive model, they do provide insight into the conformational characteristics in the RT homodimer, thereby limiting possible maturation scenarios. First, the observation that the p66 NMR spectrum is a composite of those of the folded sub-domains indicates that p66, whether monomer or homodimer, is folded in solution. Second, we found that the RNH domain is structured in the p66 homodimer, with a conformation essentially identical to that seen in the isolated RNH domain, not altered by domain-domain contacts. Third, none of our data suggests that there are different conformations in the two p66 subunits of the homodimer.

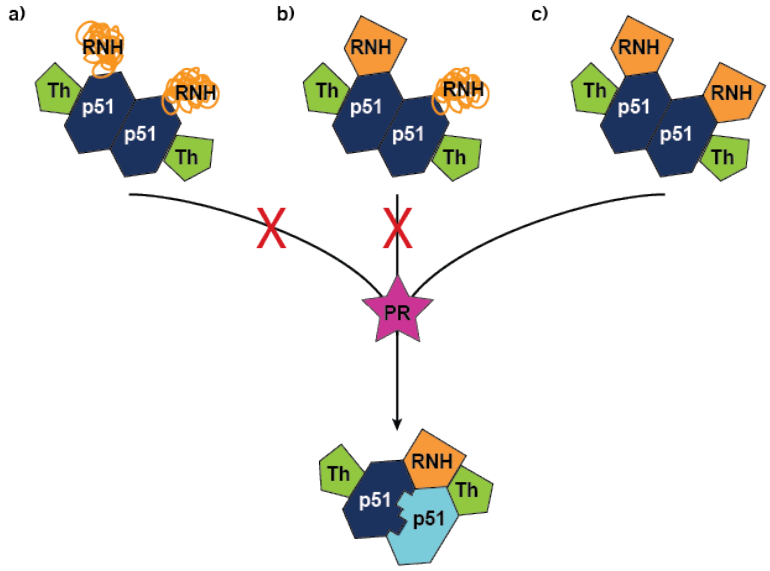

Figure 4.5 Possible RNH conformations in the p66 homodimers conformations as maturation precursors

a) the RNH domains are unfolded or disordered, b) there is conformational difference between the two RNH domains, c) both RNH domains in the two subunits are folded. Our data support c).

Our observation of a structured RNH domain in the p66 homodimer is consistent with previous biochemical results. Protease predominantly processes p66 at the p51-RNH processing site (F440-Y441), compared to other processing sites within the RNH domain in the p66 precursor (Y483-L484, N494-I495, and Y532-L533), and removal of the RNH domain does not involve multiple cleavages [23, 167]. Since our data show that the RNH in the p66 homodimer is folded, this explains why other protease cleavage sites are not used and only the authentic p51-RNH site is cleaved.

In the current study, NMR signals in finger-palm and connection domains were not assigned. Thus, our data cannot rule out a model in which the p66 homodimer exists in an asymmetric conformation. However, we can rule out a model in which the predominant (>90 %) conformation of the RNH domain is different in the two subunits of the p66 homodimer. Even if in solution a small free energy difference between the two RNH domains may exist that could give rise to two different conformations (conformational substrates) for very short times, the overall conformational equilibrium will most likely yield an average conformation over longer times. Thus, both domains will adopt the same average energy state, even if two energetically distinct conformations, separated by a low energy barrier, exist. In addition, while our manuscript was under review, Zheng, et al. published a study on HIV-RT [197]. These authors suggest, using Ile methyl signals as probes, that selective unfolding of one of the RNH domains in the p66/p66 homodimer occurs. Based on our data, we believe that major disruption of the RNH structure is unlikely, although local or temporal unfolding of a small region around the processing site cannot be ruled out.

4.5 CONCLUSION

Our results support maturation models that involve folded, rather than unfolded or disordered RNH domains (**Figure 4.5**). This is somewhat surprising since the p51-RNH cleavage site has to be processed by the protease, and thus needs to be available for binding. However, as pointed out above, if a minor conformation (<5 %) were present, this may have eluded detection, given the limited sensitivity and signal overlap in the spectra. Indeed, at present we suggest that such a minor conformation of the p51-RNH processing site can be selected or induced by the protease.

4.6 ACKNOWLEDGEMENTS

This study was supported by grants from the National Institutes of Health (R01 GM105401 to R.I. and M.A.P., and P50GM082251 to R.I. and A.M.G.), the National Science Foundation (Graduate Research Fellowship Grant, 1247842, to N.G.S.), and start-up funds from the University of Pittsburgh School of Medicine. We thank Teresa Brosenitsch for critical reading of the manuscript, Mike Delk for NMR support, Justine, H. Jou, Atticus Huberts, and Jinwoo Ahn for protein expression and purification support, and Nicolas Sluis-Cremer for the original RT DNA construct. We also thank Stuart F.J. Le Grice, Mary D. Barkley, and Robert E. London for helpful discussions.

5.0 NMR STRUCTURE OF THE HIV-1 RT THUMB SUBDOMAIN

Here, we report the solution NMR structure of the isolated thumb subdomain of HIV-1 reverse transcriptase (RT). A detailed comparison of the current structure with dozens of the highest resolution crystal structures of this domain in the context of the full-length enzyme reveals that the overall structures are very similar, with only two regions exhibiting local conformational differences. The C-terminal capping pattern of the αH helix is subtly different, and the loop connecting the αI and αJ helices in the p51 chain of the full-length p51/p66 heterodimeric RT differs from our NMR structure due to unique packing interactions in mature RT. Overall, our data show that the thumb subdomain folds independently and essentially the same in isolation as in its natural structural context. The results presented in this chapter have been adapted from: Sharaf NG[1], Brereton AE[2], Byeon I-JL[1], Karplus PA[2], Gronenborn, AM[1]. NMR structure of the HIV-1 reverse transcriptase thumb subdomain. (*submitted to the Journal of Biomolecular NMR*).

[1]Department of Structural Biology and Pittsburgh Center for HIV Protein Interactions, University of Pittsburgh School of Medicine, Pittsburgh, PA 15260.

[2]Department of Biochemistry and Biophysics, 2011 Ag & Life Sciences Bldg, Oregon State University, Corvallis, OR 97331.

5.1 INTRODUCTION

RT is an essential protein in the viral life cycle and a major drug target [198-200]. Previous crystallographic studies have shown that mature HIV-1 RT is an asymmetric heterodimer, composed of two subunits, p66 and p51. The p66 subunit contains two enzymatically active units: the polymerase and RNH domains. The polymerase domain is divided into subdomains named fingers, palm, thumb, and connection [36]. The names of the first three subdomains were inspired by the shape of the p66 subunit of this and related polymerases, in which the spatial arrangement of the thumb, fingers and palm subdomains resembles a right hand ready to clasp a piece of DNA. The p51 subunit shares the same sequence as the p66 subunit but lacks the RNH domain. The finger, palm, thumb and connection subdomains are also present in p51, although they exhibit different spatial arrangements from those in the p66 subunit (**Figure 5.1**) [37, 38, 40, 49].

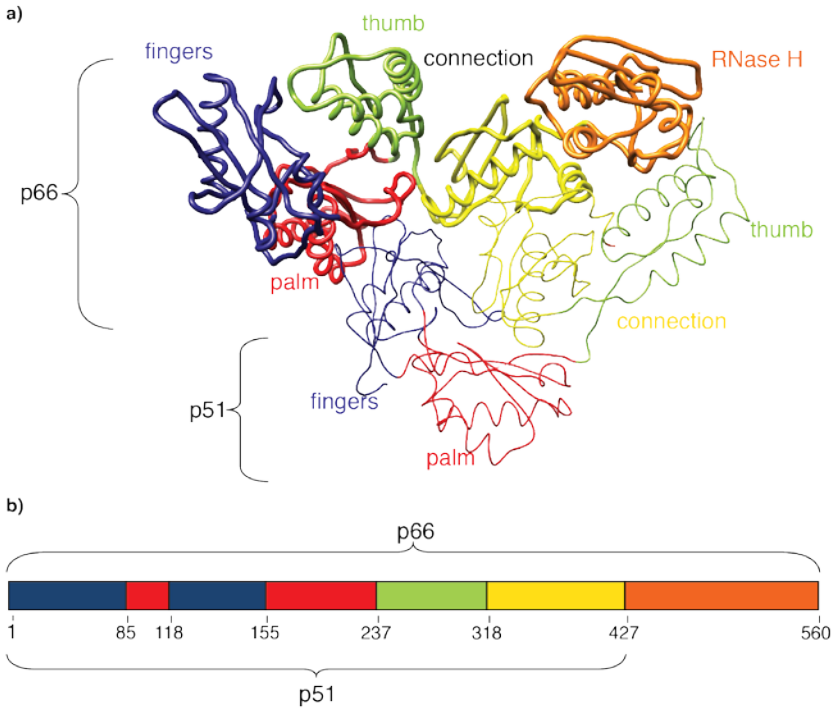

Figure 5.1 Overall HIV-1 RT structure

a) *Tube representation of apo-RT (PDB ID: 1DLO), with the p66 and p51 subunits shown in thick and thin tube representation, respectively.* ***b)*** *Schematic diagram of (sub)domain organization of HIV-1 RT. The palm, fingers, connection, thumb and RNH (sub)domains in* ***a)*** *and* ***b)*** *are colored in red, blue, yellow, green and orange, respectively.*

At present, the precise mechanism of HIV-1 RT maturation has not been elucidated. Based on model system data, a dimeric p66 immature precursor is formed first, which is subsequently cleaved at the p51-RNH processing site on one of the p66 subunits. This removes the RNH domain and generates the p51 subunit [168, 196, 201, 202]. No atomic structures of the p66:p66 immature precursor are available, although several maturation models for HIV-1 RT have been proposed [169, 170, 196, 197]. In addition, numerous biochemical/biophysical data characterizing the properties of the immature precursor exist [184, 203-208]. The p66 immature

precursor and mature HIV-1 RT exhibit similar polymerase and RNH activities, but differ in their inter-subunit affinity [209]. The mature HIV-1 RT (i.e a p66:p51 heterodimer) is a tighter dimer with a dissociation constant (K_d) of 0.23 μM, while the K_d for the p66:p66 immature precursor is 4.4 μM [210].

To structurally characterize RT maturation, we previously investigated the conformation of the p66 homodimer by solution NMR [208]. Given the protein's large size (132 kDa), NMR studies of the p66 immature precursor are challenging. However, using TROSY-type ^{1}H-^{15}N HSQC spectroscopy, it was possible to assess the structures of the (sub)domains within the p66 immature precursor and the isolated (sub)domains of HIV-1 RT. ^{1}H-^{15}N HSQC resonances of the isolated HIV-1 thumb subdomain superimposed well with the equivalent resonances in the ^{1}H-^{15}N HSQC spectrum of the homodimer, suggesting that the thumb domains in the p66 immature precursor exhibit very similar conformations as the isolated thumb domain [208]. Here, we extend this work by reporting the solution structure of the isolated HIV-1 RT thumb subdomain, and showing that it too is very similar to the crystal structures of the thumb domains in the mature RT.

5.2 EXPERIMENTAL PROCEDURES

5.2.1 NMR spectroscopy

Uniformly ^{15}N- and ^{13}C,^{15}N-labeled proteins were buffer exchanged into 25 mM sodium phosphate, 100 mM NaCl, 10% v/v D_2O, pH 6.8 in an Amicon Ultra concentrator (EMD Millipore) to a final volume of 350 μL and final protein concentration of 1.0 mM. All NMR

spectra were acquired at 30 °C on Bruker AVANCE600 and AVANCE700 spectrometers, equipped with 5 mm triple resonance, Z-axis gradient cryoprobes (Bruker Biospin, Billerica, MA). Backbone and side chain resonance assignments were carried out using two-dimensional (2D) ^{1}H-^{15}N HSQC, three-dimensional (3D) HN(CO)CACB, HNCACB, H(CCCO)NH, C(CCO)NH, HCCH-TOCSY spectra [86]. Distance restraints were derived from 3D simultaneous ^{13}C- and ^{15}N-edited NOESY spectra [211], using a mixing time of 0.15 s. All NMR data were processed with TOPSPIN 2.1 or 3.1 (Bruker) and NMRPipe [187], and analyzed with Collaborative Computing Project for NMR (CCPN) analysis program [212].

5.2.2 NMR structure calculation

Structure calculations were performed using the anneal.py protocol in XPLOR-NIH [213]. An iterative approach with extensive manual cross-checking of all distance restraints against the NOESY data and intermediate structures was employed using CCPN analysis. The final number of the NMR-derived restraints was 2,782, with 2,620 NOE distances, 46 H-bond distances identified from NOE patterns for α-helices and β-sheets, and 116 ϕ and ψ backbone torsion angles from TALOS+ calculations [214]. Two hundred and fifty six structures were generated and the 30 lowest energy structures were selected and analyzed using PROCHECK-NMR [215] and MolProbity [216]. Atomic coordinates of the structures have been deposited in the Protein Data Bank with accession code 5T82, and chemical shift assignments have been deposited in the Biological Magnetic Resonance Data Bank accession number 30171. Structures were visualized with MOLMOL [217] and VMD [218]. Structural figures were generated using UCSF Chimera [219] and the PyMOL Molecular Graphics System, Version 1.8 [220].

5.2.3 Ensemblator comparisons of NMR and crystal structure ensembles

A representative set of high-resolution RT crystal structures, refined at a resolution of 2.4 Å or better, was selected from all deposited structures. This yielded the following set of 28 PDB entries: 1RTJ, 1S1T, 2OPS, 2RKI, 2YKN, 2YNF, 2YNG, 2ZD1, 3BGR, 3DLG, 3DLK, 3LAK, 3LP1, 3MEC, 3MEE, 3QIP, 3T1A, 4DG1, 4I2P, 4ID5, 4IDK, 4IFV, 4IFY, 4IG3, 4KFB, 4KO0, 4KV8, and 5D3G. Analyses were carried out using Ensemblator version 3 (https://github.com/atomoton/ensemblator; Brereton and Karplus, unpublished), a greatly enhanced version of the recently described general ensemble-ensemble comparison program [221]. The Ensemblator "Prepare" stage was used to combine the 30-member NMR ensemble with all the crystal structures to generate a file that contained all of the backbone and side-chain atoms that were in common between the NMR models and the 28 individual p51 and p66 chains from the X-ray structures. For all Ensemblator "Analysis" stage runs, a distance cutoff of 2 Å was used to define "core" atoms; with this cutoff, 26.9 % of the atoms in the ensemble qualified for the common-core, which was used to guide the global best-fit. Based on these best-fit models, two automated clustering methods inherent to the program were used, as well as a "manual" approach, in which the NMR models, the p51 chains and the p66 chains were separated into three groups.

For each comparison, the standard Ensemblator output [221] included residue-level plots of the global and local comparisons for each pair of groups. In order to identify the most significant/consistent regions of similarity and any differences between a given pair of grouped structures, a residue-level "silhouette index" was calculated with Ensemblator v3. This index combines both global and local comparison information. For each of the two groups in a given M,N pair of grouped structures, every atom's global silhouette score is calculated as the mean

pairwise distance between the groups minus the mean pairwise distance within the group, divided by the larger of the two values: ($<d_{inter}>-<d_{intra}>$)/max($<d_{intra}>$,$<d_{inter}>$). The silhouette scores for each atom are averaged across the two groups, and a residue-based value is obtained by averaging the values for the N, CA, C', and O atoms of each residue. A second silhouette score for comparing local backbone conformations is similarly calculated for each residue based on the "locally-overlaid dipeptide residual" (LODR) distances [221]. The final "silhouette index" for a residue is the average of the global and local silhouette scores. The level of detectable difference between the groups increases with the index as it goes from near 0 to 1. For indices 0.4 – 0.6, we have considered the groups to be neither notably similar nor different; within this range, more fine clustering may permit the identification of subgroups with some differences.

5.3 RESULTS AND DISCUSSION

5.3.1 Solution structure of the HIV-1 RT thumb domain

The ^{1}H-^{15}N HSQC spectrum of the thumb subdomain exhibits well-dispersed resonances (**Figure 5.2a**), indicative of a well-folded structure. Near complete (> 95 %, backbone and side chain) NMR assignments were obtained. A superposition of the backbone atoms (N, Cα, and C') of the final 30-member ensemble is shown in **Figure 5.2b**. A summary of structural statistics is provided in **Table 5.1**, demonstrating that the core domain structure is well defined with an average atomic root mean square deviation (RMSD) of 0.45 ± 0.05 Å and 0.97 ± 0.08 Å for the backbone and all heavy atoms (residues 246-314). **Figure 5.2c** displays the lowest energy

structure in ribbon representation, illustrating the architecture of the protein. As seen in RT crystal structures, the fold of the thumb domain consists of three α-helices, commonly named αH, αI, and, αJ [36] that are linked together by loop regions that lack regular secondary structure. In the solution structure of the isolated thumb domain determined here, the core of the structure similarly comprises three α-helices, with αH (254-270), αI (278-285), and αJ (297-311). The greatest backbone variation in the ensemble occurs at the two termini and the 285-295 loop between αI and αJ.

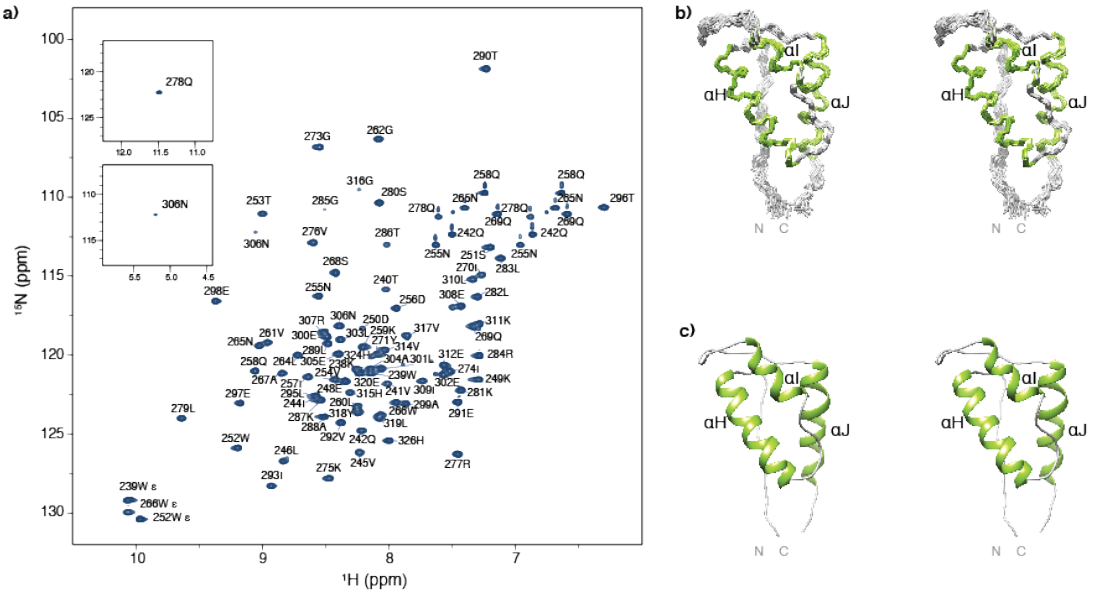

Figure 5.2 Assignments and solution structure of the thumb subdomain

a) 600 MHz ^{1}H-^{15}N HSQC spectrum at of the thumb subdomain (1.0 mM protein in 25 mM sodium phosphate, 100 mM NaCl, pH 6.8). The two resonances (Q278 and N306) that are located outside the displayed spectral range are shown in insets. b) Stereoview of the final 30 conformer ensemble (N, Cα, and C', residues 240-315). Regions of helical structure are colored in green and the remainder of the structure in grey. c) Ribbon representation of the lowest energy structure using the same color scheme as in b.

Table 5.1 Table 1. Statistics for the final 30 conformer ensemble of the thumb subdomain of RT

Number of NOE distance restraints	
Intra-residue (i-j=0)	1091
Sequential (|i–j|=1)	617
Medium range (2≤|i-j|≤4)	456
Long range (|i-j|≥5)	456
Total	2620
Number of hydrogen bond restraints	46
Number of dihedral angle restraints	
ϕ	59
ψ	57
Total	116
Structural Quality	
Violations[a]	
Distances restraints (Å)	0.029 ± 0.001
Dihedral angles restraints (°)	0.431 ± 0.091
Deviation from idealized covalent geometry	
Bond lengths (Å)	0.003 ± 0.000
Bond Angles (°)	0.478± 0.010
Improper torsions (°)	0.302 ± 0.012
Average RMSD of atomic coordinates (Å)[b]	
Backbone heavy atoms	0.45 ± 0.05
All heavy atoms	0.97 ± 0.08
Ramachandran plot analysis (%)[c]	
Favored regions	77.3 ± 3.0
Allowed regions	95.8 ± 1.9

[a] No individual member of the ensemble exhibited distance violations > 0.5 Å or dihedral angle violations >5°.

[b] The average RMSD of atomic coordinates for residues 246-314 was calculated for individual structures with respect to the mean structure. The terminal regions (residues 237-245 and 315-326) were excluded from the statistics.

[c] Statistics were calculated using MolProbity for residues 246-314; of the 4.2% of residues in disallowed' regions, all are near the allowed/disallowed borders

5.3.2 Comparisons with crystal structures of the thumb domain in the context of HIV-1 RT

The structure of HIV-1 RT thumb domain in the context of the full-length protein was first determined using X-ray crystallography [36] (**Figure 5.1a**) and since then has been seen in over 100 crystal structures deposited in the PDB. In the p66 subunit, the thumb subdomain interacts mostly with the connection domain in the same subunit and is poised to make extensive

interactions with DNA. In the p51 chain, the thumb subdomain similarly interacts with its own connection domain but also packs against the RNH domain of the p66 subunit (**Figure 5.1**). Although the thumb subdomains in p66 and p51 are found in different positions in the mature enzyme, their structures are quite similar, and both contribute to binding and positioning of nucleic acid substrates [37, 39, 222]. In the crystal structure of HIV-1 RT bound to DNA, αH of p66 interacts with the sugar-phosphate backbone of the primer strand, while the antiparallel αI interacts with the template strand [39]. In p51, in which the thumb subdomain is located next to the RNH domain of p66, the thumb domain forms the "floor" of the nucleic acid binding cleft, contributing to RNA/DNA binding [37, 39, 57, 222]. Amino acid changes in the αH and αI helices were shown to affect DNA binding, DNA synthesis, and frameshift fidelity [223-227]. αH and αI are part of the helix-turn-helix segment, termed the "helix clamp" motif, with similar motifs found in many eukaryotic, prokaryotic and viral nucleic acid polymerases [227].

In terms of overall chain-fold, the isolated domain NMR solution structure presented here (**Figure 5.2b,c**) is similar to the crystallographic subdomain structures in over 100 available crystal structures of the p51 and p66 chains of heterodimeric HIV-1 RT (**Figure 5.3a,b**). Within the complete set of RT crystal structures, many structures are of moderate resolution (3 Å or lower), and, in such structures, the conformational details are less reliable; we, therefore, selected a high-resolution subset of 28 RT crystal structures, determined at 2.4 Å resolution or better. This set was subjected to analysis by the Ensemblator [221] to identify systematic and significant differences between the NMR thumb domain structure and the p51 and p66 X-ray thumb domain structures. Automatic clustering by the Ensemblator separated the full set of models into distinct groups, also containing exclusively the NMR structure or the p51 or p66 thumb domain crystal structures. Therefore, we carried out our final analyses by manually defining these as distinct

groups. The key Ensemblator output for a given comparison of two groups consists of a pair of plots that reflect the global and local conformational differences, respectively. In each plot, the intra-group variations are compared with the inter-group variation and the closest approach between the groups (**Figure 5.4**, upper and middle panels). The regions of greatest systematic difference are characterized by high silhouette index values (**Figure 5.4**, lower panels), which contain information from both the global and local comparisons.

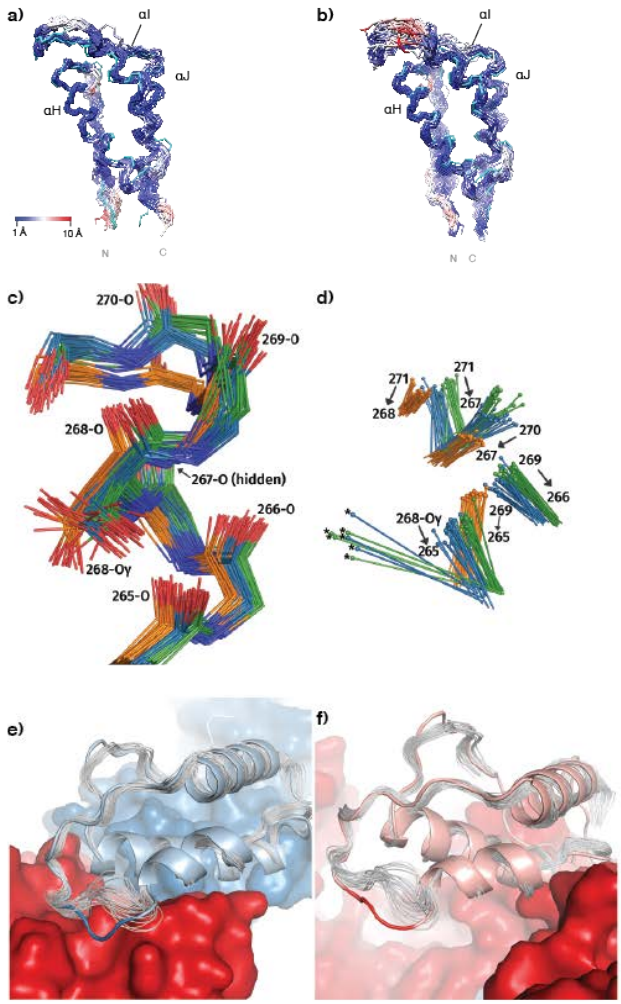

Figure 5.3 Comparison between the NMR structure of the isolated thumb domain and the thumb domain in the p51 and p66 chains in X-ray structures of the heterodimeric HIV-1 RT

a) Backbone superposition of p51 thumb domain X-ray structures and the mean NMR structure (cyan), with atoms in the X-ray structures colored from blue to red with increasing RMSD values. b) Equivalent superposition as in a), but for the p66 thumb domain X-ray structures. c) Detailed view of residues 265-271, including the side chain of Ser268, for conformers in the NMR ensemble (orange carbons), p51 chains (blue carbons) and p66 chains (green carbons). d) Hydrogen bonds between the H-atom of the donor (sphere) and the backbone oxygen acceptor in the NMR (orange), p51 (blue), and p66 (green) X-ray structures of the region shown in panel c). e) Ribbon diagram of the p51 thumb domain (light blue) from a representative RT crystal structure (PDB Code 4IFY), with residues 285-287 depicted in blue. The

remainder of p51 and p66 is shown in blue and red, respectively. The NMR ensemble (grey) differs from the X-ray structure around residue 287, where the p51 thumb subdomain contacts p66. F) Ribbon diagram of the p66 thumb domain (pink) in a representative RT crystal structure (PDB Code 4IFY), highlighting the solvent exposed position of residues 285-287 (red coil) . The remainder of the p66 chain is shown in red surface representation. In the NMR ensemble (grey), a similar local conformation is seen for residues 285-287.

In addition to some differences that occur near the N- and C-terminal residues extending from the core of the domain (<254 and >310), the silhouette indices reveal two areas of consistent differences (> ~0.6 Å) between the NMR and crystal structures (**Figure 5.4a,b** lower panels). One is near residue 270 and applies to the NMR group as compared to both the p51 and p66 thumb domain groups. The second is around residues 285-295 and applies to the NMR ensemble as compared to the p51 group, but not the p66 group (**Figure 5.4**). In both of these regions, no individual member of the NMR ensemble is closer than 1 Å to any member in the crystal structure ensembles, either globally or locally (red traces in **Figure 5.4a,b** upper and middle panels). In both cases, these emerged as real differences between the conformation of the isolated thumb subdomain and in the mature p66:p51 heterodimer.

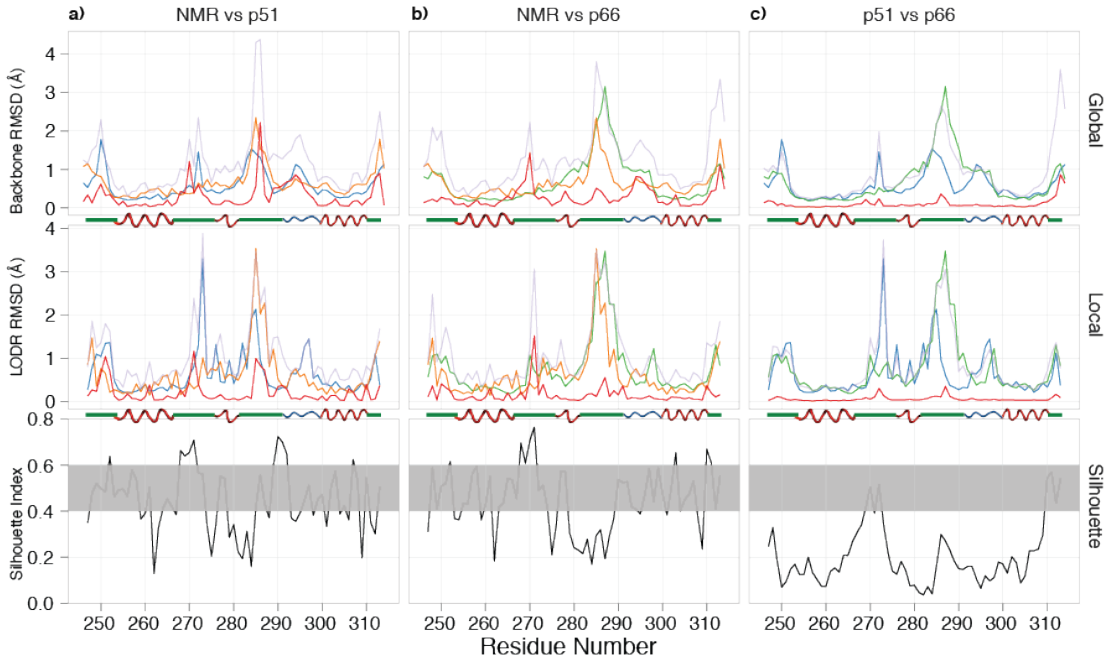

Figure 5.4 Comparison between the NMR ensemble and the collection of 28 p51 and p66 thumb domain crystal structures

a) *NMR vs. p51.* ***b)*** *NMR vs. p66.* ***c)*** *p51vs. p66. For each pair of groups, global (upper panels), and local (middle panels) comparisons are shown, along with a Silhouette Index (lower panels; black trace) plot that identifies the regions with significant differences (above the grey strip) and similarities (below the grey strip). In the global and local comparison plots, the colors are as follows: NMR average backbone intra-group pairwise RMSD (orange), p51 average backbone intra-group pairwise RMSD (blue), and p66 average backbone intra-group pairwise RMSD (green) are shown along with average backbone inter-group pairwise RMSDs (pale purple) and closest approach distances (red). A secondary structure diagram (based on PDB entry 1RTJ [50]) indicates α-helical (red), P_{II}-helical (blue), and other (green) segments.*

The difference near residue 270 involves the last turn and C-terminal capping of the αH helix. In this segment, in all of the NMR conformers, the transition from an α- to a 3_{10}-helix

occurs at residue 267, with hydrogen bonds between 269-N ··· 265-O, 270-N ··· 267-O and 271-N ··· 268-O. In contrast, in all of the crystal structures this transition occurs one residue earlier and the structures contain H-bonds between with 269-N ··· 266-O and 271-N ··· 267-O, along with a side chain-backbone H-bond from Ser268 O_γ to 265-O (Fig 3c,d). We carefully examined this difference and ascertained that indeed it is real; the observed NMR NOE pattern is incompatible with the 271-N ··· 267-O H-bond seen in the crystal structure, although the cause for this discrepancy is not clear. Inspection of this region in the p51 and p66 chains in the crystal structures shows diverse (i.e. subunit and crystal form-dependent) packing interactions with either distant parts of the chain or across crystal contacts. For example, the structural contexts of the p66 and p51 thumbs are distinct: the p66 helical cap packs against and makes H-bonds with the connection domain near residue 345, but this part of the p51 helix is more exposed while the sidechain of Ile270 packs deeply into a hydrophobic pocket that also includes Phe346 and Trp426. Because these interactions are not uniform and present in all structures, it is hard to ascertain whether and how they may influence the helix capping pattern. One set of interactions that is common between p66 and p51 involves side chain packing of Tyr271 against residues 310 through 314. These interactions are also similar in the isolated domain NMR structure, thus do not appear responsible for the difference between the X-ray and NMR structures.

The difference between the p51 and p66 thumb domains near residue 285 can easily be rationalized by the unique context of the p51 chain in the heterodimeric RT. In p51, the 285-295 loop between αI and αJ packs against the surface of p66 (**Figure 5.3e**), with a consistent hydrogen bond between the Lys287 backbone oxygen and Tyr441-O_ζH of the p66 chain. By contrast, in the p66 thumb domain, this loop is fully exposed to solvent in some crystal forms (e.g. **Figure 5.3f**) or involved in crystal contacts in others. This results in a number of diverse

conformations that vary over ~3 Å among the different crystal structures. This spread is larger than the ~2.5 Å backbone intra-group RMSD observed in the NMR ensemble (**Figure 5.4b** upper panel). Although the local conformation in this area is somewhat similar between the NMR ensemble and the p66 crystal structures, there appears to be a global shift of ~1 Å (**Figure 5.3f**). Interestingly, there is a low silhouette index near the 285 region, when p51 and p66 crystal structures are compared (Figure 4c), indicating that no large consistent difference is present; this occurs because among the broad conformations sampled in the p66 chains, some structures adopt a conformation similar to that seen in p51.

5.4 CONCLUSION

We previously suggested that the structure of the isolated HIV-1 thumb subdomain resembles that of the thumb domains in the p66 immature precursor, based on a qualitative comparison of their ^{1}H-^{15}N HSQC spectra [208]. Here, we directly confirm that the NMR structure of the isolated HIV-1 thumb subdomain is very similar to the thumb domain in X-ray structures of mature RT, albeit with two regions of interesting local conformational difference. One pertains to the C-terminal capping pattern of the αH helix, with no apparent cause. The other involves the loop conformation between the αI and αJ helices in the p51 chains. This most likely originates from unique packing interactions of the p51 thumb domain in mature RT that are not present in the isolated domain. Taken together, we show that the thumb subdomains in the mature RT and the p66 immature precursor are independent units that can fold autonomously and exhibit very similar structures, whether in isolation or present in its two different natural structural contexts

(p51 or p66). Our data also underscore the well-known fact that surface regions are malleable and are influenced by context.

5.5 ACKNOWLEDGEMENTS

This work was supported by National Institutes of Health grants P50GM082251 (to AMG) and R01GM083136 (to PAK). N.G.S was the recipient of a Graduate Research Fellowship, 1247842, from the National Science Foundation.

6.0 THE CONFORMATIONAL PLASTICITY OF HIV-1 RT

HIV-1 reverse transcriptase (RT) is a major drug target in the treatment of HIV-1 infection. RT inhibitors currently in use include non-nucleoside, allosteric RT inhibitors (NNRTIs), which bind to a hydrophobic pocket, distinct from enzyme's active site. We investigated RT-NNRTI interactions by solution ^{19}F NMR, using singly ^{19}F labeled RT proteins. Comparison of ^{19}F chemical shifts of fluorinated RT and drug-resistant variants revealed that the fluorine resonance is a sensitive probe for identifying mutation-induced changes in the enzyme. Our data show that in the unliganded enzyme, the NNRTI-binding pocket is highly plastic and not locked into a single conformation. Upon inhibitor binding, the binding pocket rigidifies. In the inhibitor-bound state, the ^{19}F signal of RT is similar to that of drug-resistant mutant enzymes, distinct from what is observe for the free state. Our results demonstrate the power of ^{19}F NMR spectroscopy to characterize conformational properties using selectively ^{19}F labeled protein. The results presented in this chapter have been adapted from: Sharaf NG, Ishima R, Gronenborn AM (2016) Conformational Plasticity of the NNRTI-Binding Pocket in HIV-1 Reverse Transcriptase: A Fluorine Nuclear Magnetic Resonance Study. Biochemistry-Us 55:3864–3873.

6.1 INTRODUCTION

HIV-1 reverse transcriptase (RT) is an essential enzyme in the HIV-1 lifecycle and a major drug target in the treatment of HIV-1 infection. Current FDA approved RT inhibitors are effective, but continuous treatment can lead to the emergence of drug resistant strains [3]. Understanding RT, its structure, and the mechanism of inhibitor action, is important for the development of novel inhibitors with more favorable resistance profiles. A large number of crystal structures of RT are available (wild-type and mutants), providing valuable information on the protein's conformations as well as drug interactions [36, 38, 46, 47, 49, 54, 55, 228]. Crystallographic studies have shown that RT is an asymmetric heterodimer that comprises two subunits p66 and p51. The p66 subunit contains two domains, a polymerase, and RNH domain. The p51 subunit is identical in amino acid sequence to p66, apart from lacking the C-terminal RNH domain. The polymerase domain of each subunit is further subdivided into fingers, palm, thumb, and connection subdomains [36]. In the overall dimeric RT structure, the subdomains in the p51 and p66 subunits exhibit different relative orientations (**Figure 6.1a**).

Although highly effective as RT inhibitors and the first drugs to treat HIV-1 infection, nucleoside/nucleotide RT inhibitors, which act as chain terminators in the enzymatic reaction, are associated with numerous side effects. Therefore, non-competitive RT inhibitors were developed and have been in the clinic for almost 20 years [5, 25, 198-200, 229]. These NNRTIs include NVP, EFV, ETR, and RPV. Although chemically diverse, they all bind to the same pocket, distinct from the polymerase active site, and inhibit RT allosterically [59, 199]. A comparison of the crystal structures of apo-RT and RT in the presence of NNRTIs reveals significant structural changes upon NNRTI binding. Apo-RT adopts a “closed” conformation, in which the p66 thumb subdomain folds down onto the fingers subdomain. In contrast, in the

presence of NNRTIs, RT adopts an “open” conformation, in which the p66 thumb domain is ~30 Å away from the fingers subdomain (**Figure 6.1b**). Local conformational differences are also seen in the NNRTI-binding pocket in the p66 subunit. This pocket is not present in apo-RT, where the Y181 and Y188 side chains fill most of the cavity, which is occupied by the NNRTI in the NNRTI/RT complex (**Figure 6.1c**)[36, 38, 46, 49].

Crystal structures are invaluable for pinpointing structural details of enzyme-inhibitor and substrate interactions, however, studies by other methods can offer complementary information. For RT, only a limited number of investigations in the absence of nucleic acid substrates have been reported, including EPR experiments and hydrogen exchange mass spectrometry (HXMS)[230, 231]. Also, a few solution NMR studies, using [methyl-^{13}C]-methionine or isoleucine labeled RT have been reported [177, 207]. In addition, several computational studies have been conducted to characterize RT dynamics and the effects of NNRTI binding [232-239]. Yet, a general consensus on the mechanistic basis for NNRTI inhibition of RT has not been reached [5, 36, 50, 240].

Here, we used ^{19}F solution NMR to study RT in solution by incorporating a single fluorine probe into the enzyme. Single site labeling prevents resonance overlap and enables simple and fast 1D NMR experiments. The fluorine nucleus was selected since it possesses a high gryomagnetic ratio, which results in excellent sensitivity (83% of ^{1}H). In addition the ^{19}F shielding is dominated by a large paramagnetic term, which makes it exquisitely sensitive to its local environment (the ^{9}F chemical shift range is ~100 fold larger than that of ^{1}H) [129, 130, 153].

In particular, we aimed to gain insight into the characteristic dynamics of the NNRTI-binding site, mechanism of action of NNRTIs, and the effects of drug resistant mutations. For

this study, RT was labeled site-specifically with 4-trifluoromethylphenylalanine (tfmF) at positions 127, 146 and 181, producing three singly labeled RT proteins named RT127tfmF, RT146tfmF, and RT181tfmF, respectively. Comparisons of the ^{19}F spectra of RT127tfmF, RT146tfmF, and RT181tfmF showed that distinctly different chemical shifts are observed for the trifluoromethyl group, demonstrating that the ^{19}F probes in each RT protein are in distinct environments. Furthermore, linewidth analyses of these spectra suggest that the NNRTI-binding site is highly plastic in the ligand-free enzyme. Mutant, drug resistant proteins V108I, K103N, and E138K all modulate the conformational plasticity or average chemical environment of the NNRTI-binding site, with the K103N mutation producing the most prominent effect. In the presence of NVP, EFV, ETR, and RPV, the conformational plasticity of RT at the NNRTI-binding site is reduced, and the chemical shift of the NNRTI-bound signals depends on the identity of the inhibitor, regardless of the presence of drug resistant mutations. These data show that ^{19}F NMR can be used as an effective tool for examining NNRTI -RT interactions.

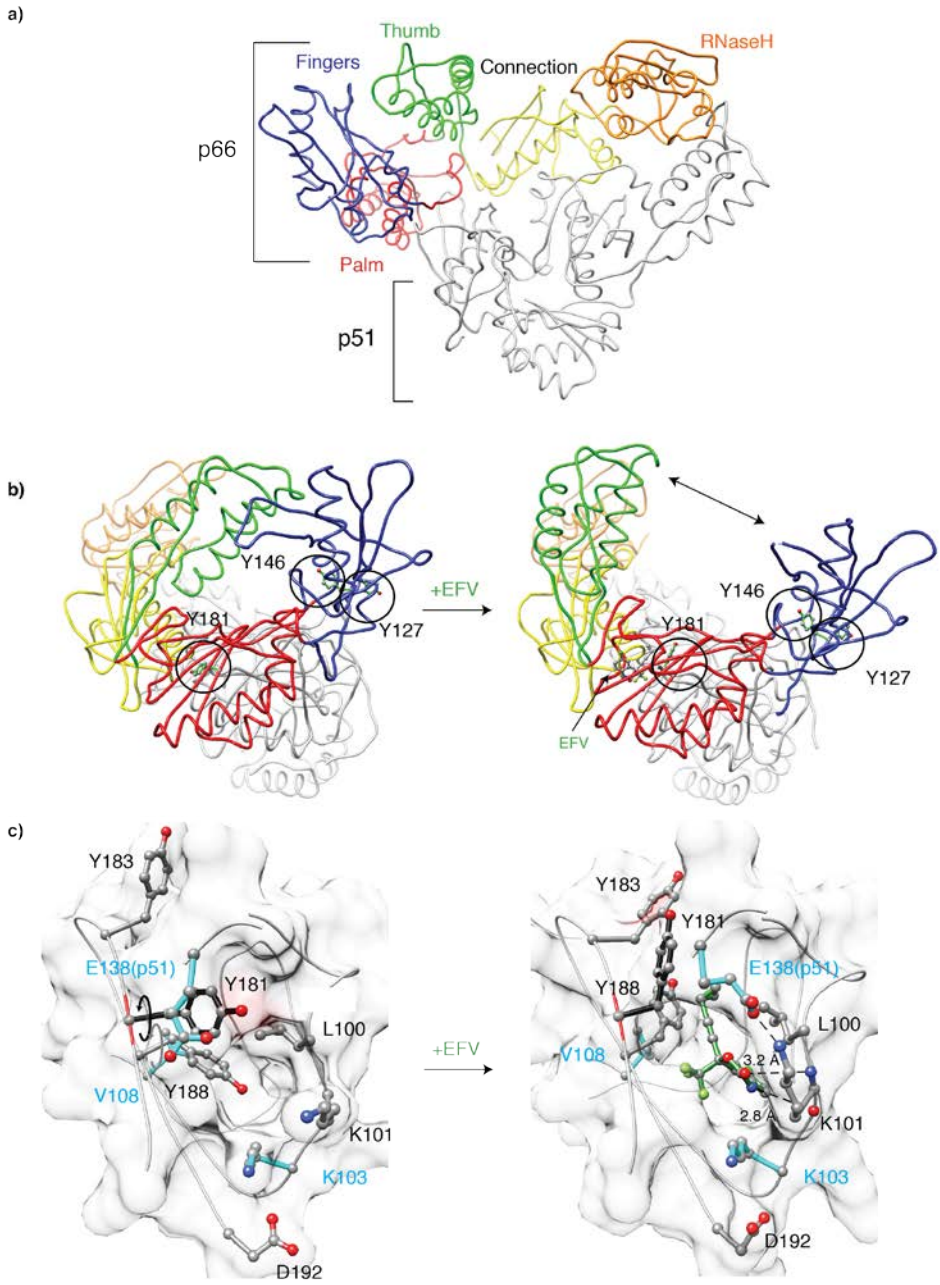

Figure 6.1 General description of RT structure, and comparison of apo and EFV-bound crystal structures of RT

a) Tube representation of apo-RT (PDB: 1DLO), with the fingers, palm, thumb, connection, and RNH domains in the p66 subunit colored in blue, pink, green, yellow and orange, respectively. The p51 subunit is colored grey. a) Structural differences between apo-RT (left, PDB: 1DLO) and EFV-bound RT (right,

PDB: 1FK9 [46]). A large conformational change, including the separation of the thumb and fingers domains (indicated by the arrow), is seen in the drug-bound structure. Tyrosine residues 127, 146 and 181 are depicted in ball and stick representation and encircled. c) Details of the binding site in apo RT and the EFV-bound RT complex, illustrating the rotation of the Y181 (black arrow) and Y188 (grey arrow) side chains out of the binding pocket. The bound EFV molecule is shown in green and pertinent distances between the benzoxazin-2-one and the backbone carbonyl oxygen of K101 (2.8 Å), and the carbonyl group of the bexzoxazine-2-one and the backbone nitrogen of atom K101 (3.2 Å) are indicated.

6.2 EXPER IMENTAL PROCED URES

6.2.1 Proteins and reagents

All proteins were prepared as described in **Chapter 3**, see **Table 3.1**. An expanded description of each RT protein is also included below.

The RT127tfmF, RT146tfmF, and RT181tfmF proteins contain a single tfmF in the p66 subunit of RT at positions 127, 146 and 181, respectively. RT181tfmF-V108I and RT181tfmF-K103N proteins contain a singl e tfmF residue at position 181 in the p66 subunit, and amino acid changes (V108I or K103N) in both, the p51 and p66 subunits. The RT181tfmF-E138K(p51) proteins contain a single tfmF at position 181 in the p66 subunit, and the E138K amino acid change in the p51 subunit of RT.

EFV, NVP were kindly provided by Dr. Nicholas Sluis-Cremer. ETR and RPV were purchased from Selleckchem (Houston, TX). All NNRTIs were stored in DMSO at concentrations of 10 mM.

6.2.2 NMR experiments

Protein samples for NMR were buffer exchanged into 25 mM sodium phosphate buffer, 100 mM NaCl, 10% v/v D_2O, pH 6.8 in an Amicon Ultra concentrator (EMD Millipore, Billerica, MA) to a final volume of 350 µL. All final protein concentrations were ~35 µM. ^{19}F 1D NMR spectra with ^{1}H composite decoupling during acquisition were recorded on a 600 MHz Bruker AVANCE spectrometer, equipped with a CP TXO F/C-H-D triple-resonance z-axis gradient cryoprobe (Bruker Biospin, Billerica, MA). Spectra for the inhibitor-free proteins as well as samples containing NVP, EFV, ETR, and RPV at 1:1 and 1:5 RT: NNRTI inhibitor ratios were recorded using Topspin 3.1 (Bruker) and analyzed with MestReNova (Escondido, CA). Prior to Fourier transformation, the time-domain free-induction decays were apodized with an exponential function, using a line broadening factor of 30 Hz. Chemical shifts and linewidths were calculated using the peak deconvolution feature in MestReNova. An upper limit of uncertainty for the linewidths was qualitatively estimated by assuming that the fit error of each peak is associated with the linewidth error.

6.3 RESULTS

6.3.1 Spectra of apo-RT127tfmF, apo-RT146tfmF, and apo-RT181tfmF

^{19}F NMR spectra were recorded for RT127tfmF, RT146tfmF, and RT181tfmF, each containing a single tfmF at the indicated position in the p66 subunit of RT (**Figure 6.1b**). Different spectra were observed for the apo-proteins (**Figure 6.2Aa-c, black traces**), reflecting the distinct

environment around the tfmF side chain of each RT protein. For apo-RT127tfmF, a single resonance signal is seen at -61.5 ppm (100 ± 2 Hz; **Figure 6.2a**), while for apo-RT146tfmF, a major signal is observed at -62.0 ppm (150 ± 2 Hz), and a small additional signal is seen at -60.2 ppm (150 ± 85 Hz), slightly downfield from the major resonance (**Figure 6.2b**). The spectrum of apo-RT181tfmF exhibits a very broad signal at -60.8 ppm (500 ± 5 Hz), much wider than those of apo-RT127tfmF and apo-RT146tfmF, and a smaller signal around -61.8 ppm (**Figure 6.2c**). In each spectrum, the major signal clearly originates from the p66/p51 heterodimeric RT. The very small signal in **Figure 6.2c** may originate from the small amount (< 8 %) of p66 monomer and/or homodimer in the sample [210]. This was confirmed by comparing spectra of different RT samples (data not shown).

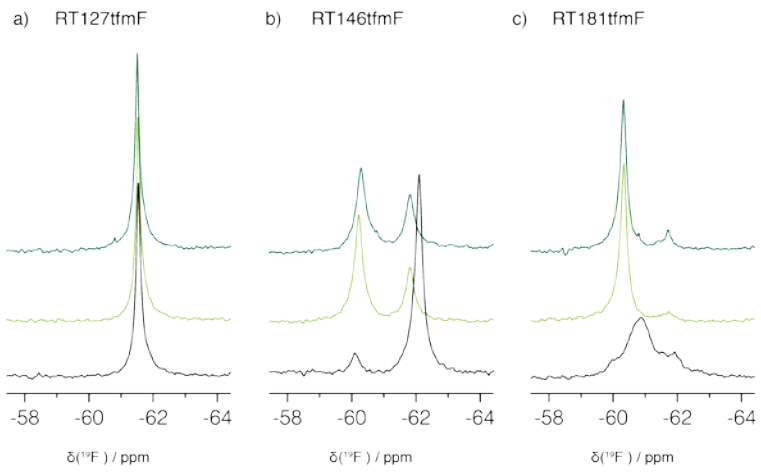

Figure 6.2 1D ^{19}F NMR spectra of RT

4-trifluoromethyl-phenylalanines substituted for tyrosine residues at several positions in the p66 subunit, in the absence (black) and presence of EFV at 1:1 and 1:5 molar ratios (light and dark green, respectively).^{19}F spectra of a) RT127tfmF, b) RT146tfmF, and c) RT181tfmF at 27°C are shown.

6.3.2 EFV binding to RT127tfmF, RT146tfmF, and RT181tfmF

The effect of EFV binding to the three RT variants, RT127tfmF, RT146tfmF, and RT181tfmF was investigated (**Figure 6.2**). At 1:1 RT:EFV molar ratio, no changes are observed for RT127tfmF (**Figure 6.2a**, light green trace). In contrast, for RT146tfmF, two new signals are observed, one very close to the one of the free protein (-62.0 ppm) and the other downfield, at -61.8 ppm (**Figure 6.2b**, light green trace). Both resonances, at -61.8 ppm and -62.0 ppm, exhibit comparable linewidths (180 ± 2 Hz). The broad asymmetric signal of RT181tfmF at -60.8 ppm (500 ± 2 Hz) disappears upon EFV binding (**Figure 6.2c**, light green trace) and a new, sharper resonance appears at -60.3 ppm (125 ± 2 Hz). Spectra were also recorded at a 1:5 RT: EFV molar ratio (**Figure 6.2**, dark green traces) to ensure saturation of the protein with ligand. For all proteins, the spectra at 1:5 are similar to spectra at 1:1 RT:EFV molar ratio, suggesting that saturation with EFV is essentially reached at the 1:1 molar ratio, consistent with the K_d value of 92 nM for EFV binding to the p66/p51 RT heterodimer [203].

These spectral data are interpreted in light of the location of the three amino acid residues in RT crystal structures [46, 52]. Residue 127 is ~35 Å away from the NNRTI binding site, located on the fingers subdomain and pointing towards the solvent (**Figure 6.1b**), and no significant changes in the position of residue 127 in the absence or presence of EFV are noted when comparing crystal structures of apo-RT and the EFV/RT complex. The resonance frequency of RT127tfmF, which is essentially not affected by EFV binding, is consistent with this observation (**Figure 6.2a**). Residue 146 is ~30 Å away from the NNRTI binding site and is also located on the fingers subdomain. However, this residue points towards the thumb domain. In the RT146tfmF protein, the ^{19}F probe at position 146 clearly senses EFV binding and splits into two resonances, at -60.2 and -61.8 ppm, in the EFV-bound form (**Figure 6.2b**). The

presence of two resonances suggests that the ^{19}F probe in RT146tfmF reports on two conformations, one of which is very similar to the free conformation, given that only a very small difference in frequency is involved, while the other one reflects a distinctly different conformation. The 181tfmF side chain is located in the NNRTI-binding pocket (**Figure 6.1b,c**) and EFV binding, by necessity, is expected to influence its conformation. The spectrum of RT181tfmF in the presence of EFV contains a substantially narrower resonance for the inhibitor bound state, which is shifted downfield compared to that of apo-RT181tfmF. The broad resonance observed in the spectrum of apo-RT181tfmF (500 ± 2 Hz) suggests that the ligand-free protein exhibits a substantial degree of conformational plasticity in the NNRTI binding site, which is reduced in the EFV/RT complex (**Figure 6.2c**), evidenced by the significantly narrower linewidth of the bound signal (125 ± 2 Hz).

6.3.3 Drug-resistant variants of RT

Three drug-resistant variants of RT were selected for investigation. V108I is associated with NVP resistance, K103N imparts NVP and EFV resistance,[241-244] and E138K is connected with therapeutic failure of ETR and RPV [245-248]. We evaluated these mutants in the context of RT181tfmF, since residue 181 resides in the NNRTI site and should report on possible effects of these mutations on the protein properties at this binding site. A superposition of the ^{19}F spectra of all three apo RT181tfmF mutants is provided in **Figure 6.3**. The linewidth of the apo-RT181tfmF-V108I signal (red, 500 ± 5 Hz) is similar to that of the apo-RT181tfmF (black, 500± 5 Hz), although slightly upfield shifted, suggesting a minimal influence of this mutation on the NNRTI binding site. In contrast, the signal of apo-RT181tfmF-K103N is narrower and downfield shifted (purple; 300 ± 15 Hz, -60.1 ppm). This suggests that in this mutant a less plastic

conformation is present in the binding site, compared to apo-RT181tfmF. The signal in the ^{19}F spectrum of apo-RT181tfmF-E138K(p51) is also broad (green, 600 ± 20 Hz) and is similar to that of apo-RT181tfmF, indicating that only small changes are introduced into the flexible binding site. Collectively, these data suggest that the local environment around the ^{19}F probe at the 181 position of RT is most prominently affected by the K103N mutation.

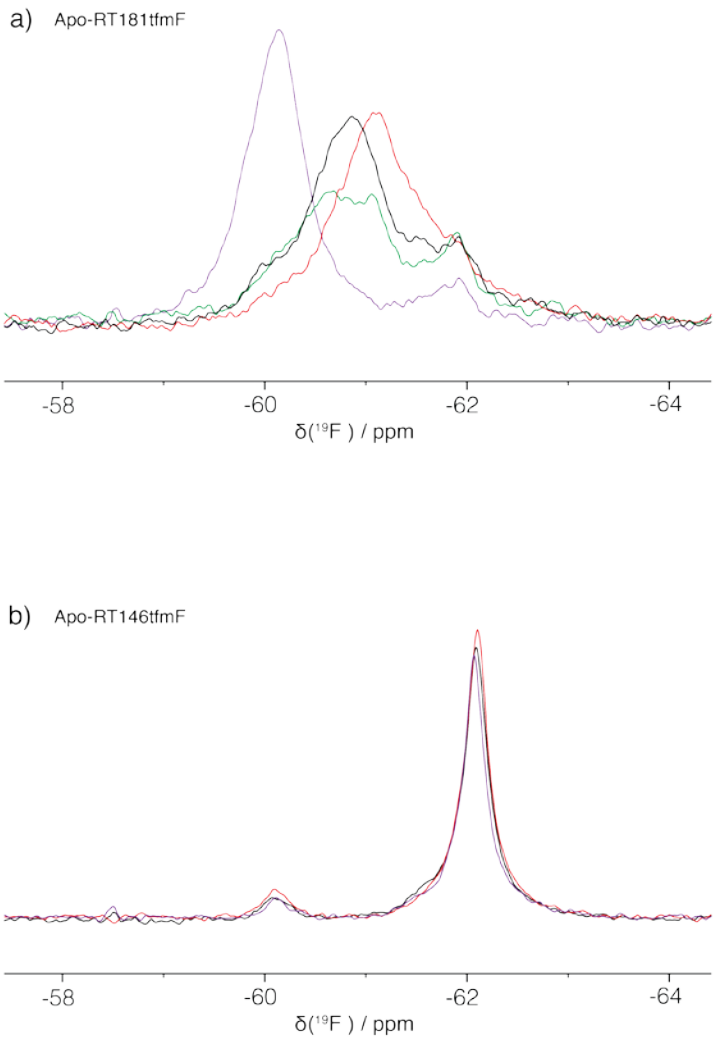

Figure 6.3 1D ^{19}F NMR spectra of RT181tfmF and several RT mutants at 27˚C.

a) Superposition of the fluorine resonances of RT181tfmF (black), RT181tfmF-V108I (red), RT181tfmF-E138K(p51) (green), and RT181tfmF-K103N (purple). b) Superposition of the fluorine resonances of RT146tfmF (black), RT146tfmF-V108I (red), and RT146tfmF-K103N (purple). All RT181tfmF and RT146tfmF variants contain amino acid changes in both the p51 and p66 domains, except for RT181tfmF-E138K in which the E138K change is only present in the p51 subunit.

6.3.4 NVP, EFV, ETR, and RPV binding to RT181tfmF and mutants associated with drug resistance

To examine the spectral perturbations of NNRTIs on the ^{19}F spectrum of RT181tfmF, we recorded spectra in the presence of NVP, ETR, and RPV, in addition to EFV (**Figure 6.4a**; note the panel of **Figure 6.2c** is included as the 2nd panel in **Figure 6.4a** for comparison). Spectra show that saturation with EFV, ETR and RPV is essentially complete at the 1:1 RT:NNRTI molar ratio, while, for NVP, much higher concentrations are needed to ensure saturation. These observations are consistent with results obtained in cell-based assays, that show the EC_{50} for NVP, 0.085 μM, is much higher than for EFV, ETR and RPV, which are 0.001, 0.002, and 0.0004 μM, respectively [58]. Importantly, in the presence of each NNRTI, the defining characteristic of each spectrum is the significantly narrower NNRTI-bound signal (**Figure 6.4a**, colored traces), compared to the inhibitor-free signal of apo-RT181tfmF (500 ± 2 Hz) (**Figure 6.4a**, black trace) (**Table 6.1**). This suggests that NVP, EFV, ETR, and RPV all reduce the conformational plasticity of the NNRTI inhibitor binding site around position 181, as reflected in the narrower linewidth of the tfmF signal.

Table 6.1 ^{19}F Resonance frequencies and linewidths[a]

	RT127tfmF		RT146tfmF		RT181tfmF		RT181tfmF-V108I		RT181tfmF-E138K(p51)		RT181tfmF-K103N	
	ppm	Hz	ppm	Hz	ppm	Hz	ppm	Hz	ppm	Hz	ppm	Hz
Apo	-61.5	100 ± 2	-60.2 -62.0	150 ± 85 150 ± 5	-60.8[b]	500 ± 5	-61.1[b]	500 ± 5	-60.8[b]	600 ± 20	-60.1	300 ± 15
EFV	-61.5	100 ± 5	-60.2 -61.8	180 ± 1 180 ± 2	-60.3	125 ± 2	-60.3	125 ± 5	-60.2	170 ± 5	-60.4	150 ± 2
ETR					-60.3	125 ± 2	-60.3	125 ± 5	-60.2	125 ± 2	-60.4	145 ± 2
RPV					-60.7	140 ± 2	-60.7	125 ± 5	-60.7	125 ± 5	-60.7	145 ± 2
NVP			-60.2 -61.9[b]	180 ± 30 200 ± 25	-59.7 -60.8[b]	180 ± 5 500 ± 10	-59.7 -61.1[b]	125 ± 10 400 ± 10	-59.7 -60.9[b]	125 ± 10 600 ± 15	-60.1[b]	300 ± 20

[a] the upper limit of uncertainty in the linewidth was qualitatively estimated as described in Materials and Methods.
[b] although the signal is not symmetric, the resonance frequency and the linewidth were extracted assuming a single peak.

We also examined how NNRTI binding affects the different drug-resistant variants of RT181tfmF. In the presence of 5-fold excess of NVP, EFV, ETR and RPV, the spectra of RT181tfmF-V108I exhibit chemical shifts and linewidths similar to those of RT181tfmF (**Figure 6.4a,b**). All bound resonances are significantly sharper (~125 Hz) than in the apo form (~500 Hz). At 1:1 RT: NVP molar ratio, as expected, the ligand-free RT signal is still present, consistent with the 2-fold larger in EC_{50} value, compared to wt-RT [249]. RT181tfmF-K103N and RT181tfmF-E138K(p51) also exhibit similar changes in chemical shifts and linewidths, as seen with RT181tfmF upon NNRTI binding (**Figure 6.4c,d**; **Table 6.1**). Note that binding of NVP to RT181tfmF-K103N was not observed (**Figure 6.4c**). This agrees well with the much larger EC_{50} values reported in cell-based assays (> 1 μM) [58]. Taken together, upon NNRTI binding to RT, substantially narrower ^{19}F signals are observed for the tfmF group at position 181, demonstrating that the NNRTI-binding pocket becomes confined and looked into a more rigid conformation.

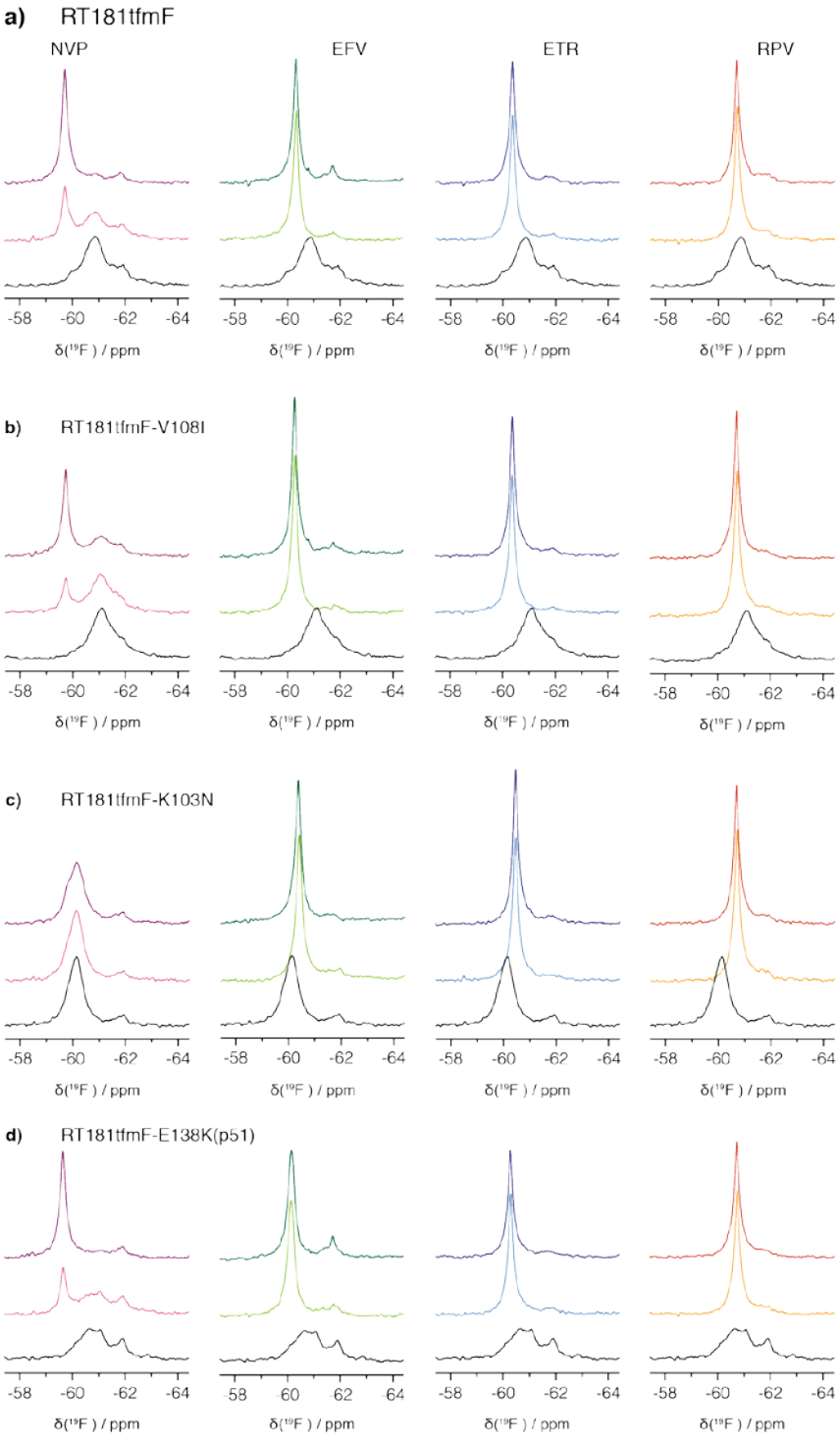
a) RT181tfmF
NVP
EFV
ETR
RPV
-58 -60 -62 -64
δ(19F) / ppm
b) RT181tfmF-V108I
c) RT181tfmF-K103N
d) RT181tfmF-E138K(p51)

Figure 6.4 1D 19F NMR spectra of RT181tfmF and several RT181tfmF mutants in the absence (black) and presence of NVP (pink), EFV (green), ETR (blue) and RPV (orange)

a) Superposition of the ^{19}F spectra of apo-RT181tfmF and the ^{19}F spectra of RT181tfmF in the presence of each NNRTI, b) Superposition of the ^{19}F spectra of apo-RT181tfmF-V108I and the ^{19}F spectra of RT181tfmF in the presence of each NNRTI, c) Superposition of the ^{19}F spectra of apo-RT181tfmF-K103N and the ^{19}F spectra of RT181tfmF-K103N in the presence of each NNRTI, d) Superposition of the ^{19}F spectra of apo-RT181tfmF-E138K(p51) and the ^{19}F spectra of RT181tfmF-E138K(p51) in the presence of each NNRTI. The ^{19}F spectra in the presence of each NNRTI at 1:1 and 1:5 molar ratios are shown in light and dark colors, respectively. Chemical formulae for each inhibitor are depicted in the individual panels.

6.3.5 NVP and EFV binding to RT146tfmF and mutants associated with drug resistance

We also recorded ^{19}F NMR spectra of RT146tfmF and its V108I and K103N mutants (**Figure 6.3b**). The spectra of these mutants in the apo form (without inhibitor) are essentially identical to each other, demonstrating that the tfmF group in position 146 is not affected by amino acid changes at positions 103 and 108 that cause drug resistance. From the ^{19}F NMR spectra of RT146tfmF and the mutants in the presence of NVP or EFV, it can be appreciated that saturation with EFV is essentially complete at 1:1 RT:EFV molar ratio (**Figure 6.5**, note that the panel of **Figure 6.2b** is included as the 2nd panel in **Figure 6.5a** for comparison). In contrast, with NVP, much higher concentrations are needed to reach saturation. As expected,[58] no NVP binding to RT146tfmF-K103N was detected. The spectra in the presence of NNRTIs exhibit two signals at approximately -62 ppm and -60 ppm, irrespective of the mutations (**Table 6.2**).

Table 6.2 ^{19}F Resonance frequencies and linewidths of mutants of RT146tfmF[a]

	RT146tfmF-V108I		RT146tfmF-K103N	
	ppm	Hz	ppm	Hz
Apo	-60.1	150 ± 30	-60.1	150 ± 30
	-62.1	150 ± 5	-62.1	150 ± 5
EFV	-60.1	180 ± 5	-60.1	180 ± 5
	-61.9[b]	300 ± 15	-61.9[b]	260 ± 5
NVP	-60.2	180 ± 10	-60.1	150 ± 30
	-61.9[b]	200 ± 25	-62.1	150 ± 5

[a] the upper limit of uncertainty in the linewidth was qualitatively estimated as described in Materials and Methods.
[b] although the signal is not symmetric, the resonance frequency and linewidth extracted assuming a single peak.

For RT146tfmF-V108I and RT146tfmF-K103N the EFV-bound signals (at 1:5 molar ratio) resonate at -62 ppm and are somewhat asymmetric, compared to that of RT146tfmF (**Figure 6.5**). However, these differences are small and may reflect minor differences in dynamics in these mutants. Overall, all available data for RT146tfmF show that no significant effect is seen in the spectra when the drug resistant mutational changes are introduced into the protein, both in the apo- and NNRTI-bound forms. This observation is consistent with the location of residue 146 in the RT structure on the fingers subdomain, ~30 Å away from the NNRTI binding site (**Figure 6.1b**).

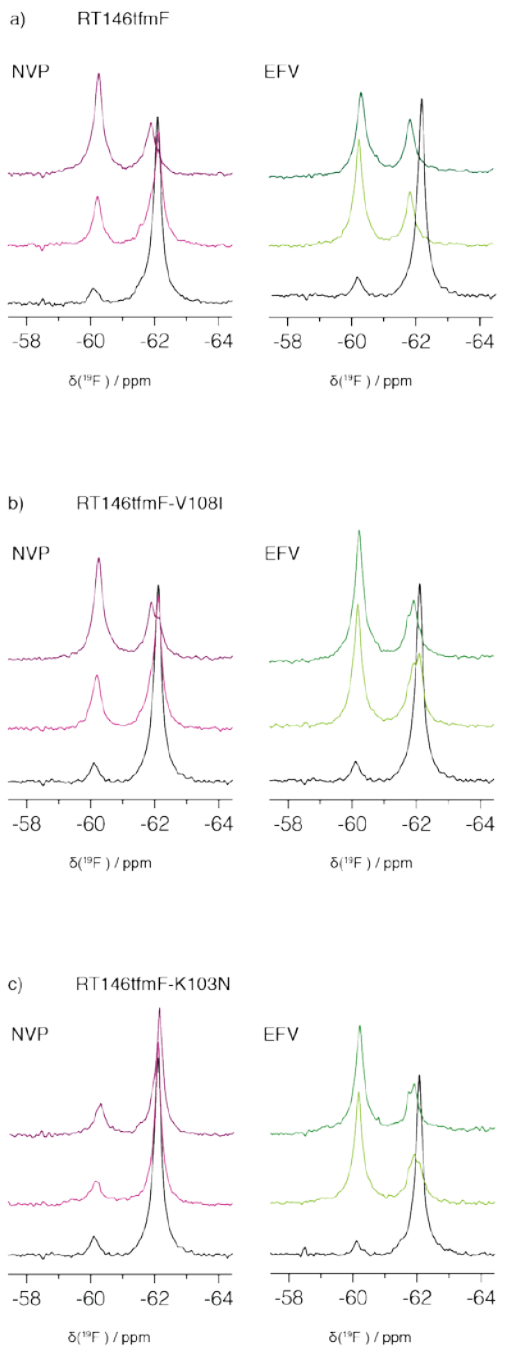

Figure 6.5 Superposition of 1D ^{19}F NMR spectra for RT146tfmF

a) RT146tfmF, b) RT146tfmF-V108I, and c) RT146tfmF-K103N, in the absence (black) and presence of NVP (pink) or EFV (green). The ^{19}F spectra in the presence of each NNRTI at 1:1 and 1:5 molar ratios are shown in light and dark colors, respectively.

6.3.6 Sensing the NNRTI interaction in RT181tfmF and RT146tfmF using the tfmF probe.

Given that residue 181 resides in the NNRTI binding site, extensive studies were carried out with different RT181tfmF variants. All chemical shifts and linewidths for these variants are summarized in **Figure 6.6a,b**. In the absence of inhibitors, the linewidths of the ligand-free signals are larger (**Figure 6.6a**) and the chemical shifts are diverse (**Figure 6.6b**). The chemical shifts of the NVP-, EFV-, ETR-, and RPV-bound RT181tfmF resonances are -59.7, -60.3, -60.3, -60.7 ppm, respectively, suggesting that different inhibitors create different chemical environments around the 181tfmF probe. Based on the linewidth data, although NNRTIs are chemically dissimilar, NVP, EFV, ETR, and RPV all seem to lock the NNRTI-binding pocket into a more rigid conformation (**Figure 6.6a**). For RT146tfmF no significant changes in resonance frequencies and linewidths in the apo- and NNRTI-bound forms among the variants are observed (**Figure 6.6c** and **6.6d**).

Furthermore, and most significantly, the inhibitor-bound chemical shifts of RT181tfmF resonances are specific for each inhibitor. Indeed, all RT variants, when complexed with NVP, exhibit the same chemical shift of -59.7 ppm. Likewise, the RPV-bound chemical shift is -60.7 ppm and EFV- and ETR-bound chemical shifts are -60.3 ± .1 ppm. These data suggest that the differences in the environment of the ^{19}F probe at position 181 in the apo-RT proteins become diminished upon inhibitor binding, resulting in a chemical environment determined by the specific inhibitor, presumably by rigidifying the NNRTI-binding pocket around the inhibitor.

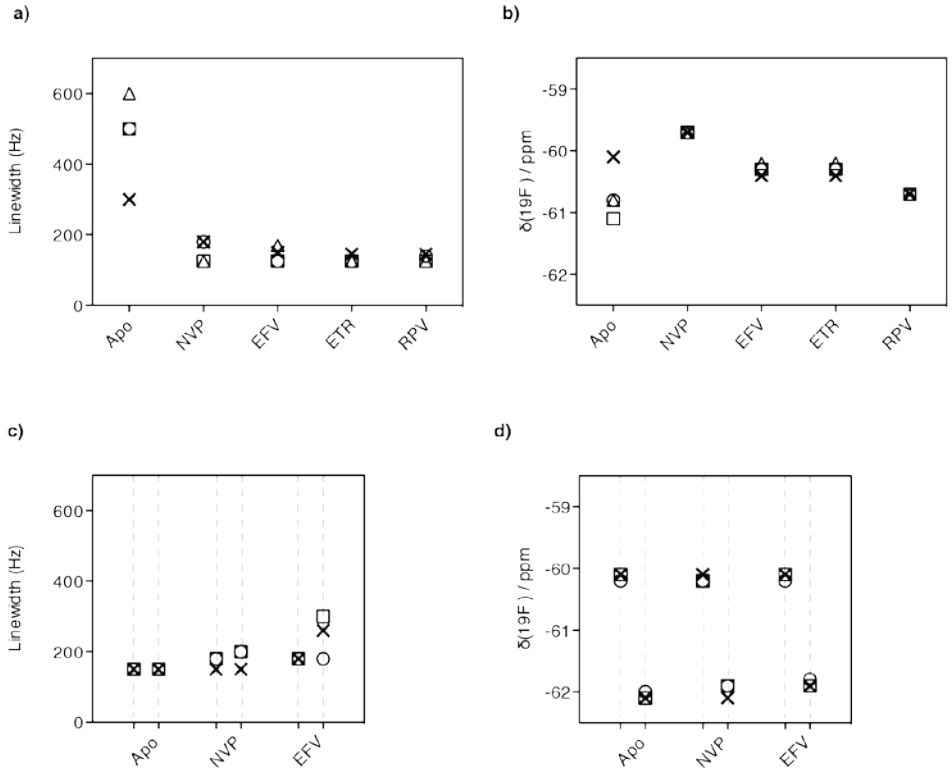

Figure 6.6 Plots of linewidths and chemical shifts of the signals in the ^{19}F spectra of RT181tfmF and RT146tfmF and the drug-resistant variants, respectively, in the absence and presence of each NNRTI. *In a) and b), plots are shown for apo- and NNRTI-bound signals of RT181tfmF (o), RT181tfmF-K103N (x), RT181tfmF-V108I (□), and RT181tfmF-E138K(p51) (△). In c) and d), plots are shown for apo- and NNRTI-bound signals of RT146tfmF (o), RT146tfmF-K103N (x), and RT146tfmF-V108I (□). Note that since the spectra of RT146tfmF and mutants thereof comprise two resonances at approximately at -62 and -60 ppm, two sets of points are contained in the plots presented in c) and d). Values were obtained from the spectra provided in Figures 4 and 5.*

6.4 DISCUSSION

^{19}F solution NMR experiments on site-specifically ^{19}F labeled RT variants were preformed to assess the proteins' behavior in the absence and presence of several NNRTIs. The three singly fluorinated RT proteins, RT127tfmF, RT146tfmF and RT181tfmF exhibit distinct spectra, reflecting the different chemical environments surrounding the ^{19}F probes. For two ^{19}F-labeled proteins, RT146tfmF and RT181tfmF, EFV binding clearly can be monitored by the probe. Interestingly, the spectrum of free RT181tfmF (**Figure 6.2c**) exhibits a very broad signal, incompatible with a single, narrow conformation of the protein, and suggests a very plastic, mobile environment that is sensed by the ^{19}F nucleus. In the crystal structures of apo-RT and NNRTI-bound RT, two different conformations of tyrosine 181 are observed, and rotation of the side chain out of the NNRTI-binding pocket in the complex is necessary to accommodate inhibitor binding [36, 38, 44, 46, 49]. However, as described in the introduction, only limited data are available for apo-RT, in contrast to an abundance of data on ligand-bound RT. Since it is essential to have access to equivalent information for both apo- and ligand-bound forms in order to evaluate conformational changes between the two states, we studied RT in solution. The spectrum of ligand-free RT suggests that the NNRTI-binding site is highly plastic in the free enzyme, and not confined to a single, narrow conformation. Thus, the solution NMR results offer different, but complementary information to the crystallographic data.

We also investigated the effect of drug-induced mutations on the chemical environment around the ^{19}F probe at position 181 (RT181tfmF-V108I, RT181tfmF-K103N and RT181tfmF-E138K(p51). Each mutation is located in a different position (**Figure 6.1c**). V108 is positioned behind residue Y188, and K103 (p51) and E138 (p51) are located at the entrance of the NNRTI-binding site. Our NMR data show that amino acid changes at these sites affect the chemical

environment of the ^{19}F probe at position 181, with the K103N change producing the greatest effects (**Figure 6.3a**). In particular, the narrow signal of RT181tfmF-K103N suggests that this mutation restricts the plasticity of the NNRTI-binding site. These results are consistent with previous studies, based on crystallographic and computational data, which suggest a "closed" form of the NNRTI-binding pocket in K103N RT [238, 250]. In contrast, ^{19}F signals of the RT181tfmF-E138K and RT181tfmF-V108I variants are significantly broader, suggesting that they possess a more plastic NNRTI-binding site. Interestingly, the latter mutations are associated with the lower degree of NNRTI resistance [246, 248, 251].

Our NMR data also revealed remarkable changes in chemical shifts and linewidths of the RT181tfmF signal upon interactions with NNRTIs. Importantly, the dynamic behavior of the 181 site is quenched upon EFV binding, consistent with previous results from HXMS experiments that showed higher protection for peptides 88-109 and 187-192 in the presence of EFV[231]. A similar effect is seen for NVP, ETR and RPV binding. We find that all the NNRTI-bound signals are narrower, indicating a significant reduction in the flexibility of the inhibitor-bound forms, compared to the apo-form (**Figure 6.6a**). Most interestingly, the chemical shifts of the different NNRTI-bound variants revealed an intriguing pattern: while the resonance frequencies of the ligand-free RT181tfmF variants, V108I, K103N and E138K(p51), are all different, compared to RT181tfmF, once a particular inhibitor is bound, these differences disappear. For all four protein complexes, essentially the same chemical shifts are noted for the inhibitor-bound proteins; thus, the NNRTI-bound shifts are characteristic for a particular inhibitor, irrespective of the presence of mutations that are associated with drug resistance. This suggests that it is the identity of the inhibitor, which ultimately determines the resonance frequencies of the tfmF probe at position 181.

While it is possible that the fluorophenylalanine substitution for tyrosine at position 181 may influence the conformation in NNRTI-binding site, this effect has to be very small, given that the binding affinities of EFV and NVP to RT181tfmF are consistent with reports in the literature [231]. In addition, the effects observed upon EFV binding when the tfmF probe resides at positions 127, 146, and 181 agree well with the distances between the ^{19}F nucleus and the NNRTI binding site. Thus, observed ^{19}F spectral changes upon NNRTI binding perhaps qualitatively reflect the native dynamics of individual sites. As described above, the ^{19}F resonances of apo RT181tfmF and mutants thereof vary (**Figure 6.3a**).

In summary, our data provide new insights into the dynamics of the NNRTI-binding site and suggest a mechanism of action for NNRTIs. Our results clearly demonstrate that the NNRTI-binding site is highly plastic in the ligand-free enzyme, and that drug resistance mutants modulate this conformational plasticity. Importantly, NVP, EFV, ETR, and RPV all reduce the dynamics of RT in the NNRTI-binding site, and the NNRTI-bound chemical shifts are determined by the identity of the inhibitor. Furthermore, the present study demonstrates that ^{19}F NMR can be used as an effective tool for examining ligand-protein interactions in cases where only small amounts of protein are available or limited solubility of protein or ligand exist.

6.5 ACKNOWLEDGMENTS

We thank Drs. Elena Matei and In-Ja Byeon for help with ^{19}F NMR spectroscopy, Mike Delk for NMR technical support, and Atticus Huberts for help with protein expression. Dr. Nicholas Sluis-Cremer is gratefully acknowledged for providing EFV and NVP, Dr. Teresa Brosenitsch

for critical reading of the manuscript, and Drs. Stuart Le Grice, Mary Barkley, Ryan Mehl and Christopher Barnes for useful discussions.

7.0 SUMMARY

To address the HIV-1 pandemic, significant efforts have been made to develop targeted therapeutics against the HIV-1 virus. As a result of multidisciplinary efforts and decades worth of research, 24 FDA approved drugs against HIV-1 are currently available. These inhibitors include NRTIs, NNRTIs, PI, FI, EI and INSTs. The use of these inhibitors as part of HAART treatment has greatly increased patient lifespan and quality of life. As a result, HIV-1 infection is now considered a manageable chronic condition, rather than a life-threatening disease [3]. Yet, despite current treatment efforts, resistance has been documented for all antiretroviral drug classes; therefore, new therapeutic regimens are needed as the virus continues to evolve and acquire resistance [2, 3, 30].

Currently half of the FDA approved drugs used to treat HIV-1 infection target RT, an enzyme that produces viral DNA using genomic viral RNA as a template. Given RT's essential role in the viral lifecycle it is still considered to be an attractive target for therapeutic intervention [4, 36]. To develop novel inhibitors with more favorable resistance profiles, it is important to gain a thorough understanding of RT, its structure, and mechanism of inhibitor action. Currently, there are over 200 X-ray structures of RT, including RT variants, and RT bound to different ligands in the PDB. These models have greatly contributed to the success of structure-based drug design and development of more potent HIV-1 drugs [3]. Although X-ray crystallography has

been invaluable for providing details of RT, studies by other methods are essential to obtain a true understanding of the protein in solution.

Solution NMR is now considered a powerful and versatile tool for protein studies. Uniquely, NMR can be used to determine protein structures and/or obtain information on a select number of NMR observable probes. Recent advances in hardware, software, pulse sequence design, processing procedures and the ability to isotopically label proteins has placed NMR side-by-side with X-ray crystallograhy. In this thesis, we used solution NMR to study the p66 immature RT precursor, which so far has resisted crystallization, and used fluorine NMR to study RT-NNRTIs interactions in solution.

Specifically, in **chapter 4**, we found that in the dimer p66 precursor, the RNH and thumb subdomains are folded and possess conformations very similar to those in mature RT. This finding suggests that maturation models which invoke a complete or predominantly unfolded RNH domain are unlikely. In **chapter 5**, we report the solution NMR structure of the isolated thumb subdomain of HIV-1 reverse transcriptase (RT). A detailed comparison of the current structure with dozens of the highest resolution crystal structures of this domain in the context of the full-length enzyme reveals that the overall structures are very similar, with only two regions exhibiting local conformational differences. Taken together, we show that the thumb subdomains in mature RT and the p66 immature precursor are independent units that can fold autonomously and exhibit very similar structures, whether in isolation or present in its two different natural structural contexts (p51 or p66). Finally, in **chapter 6** we investigated RT-NNRTI interactions by ^{19}F NMR, using singly ^{19}F labeled RT proteins. Comparison of ^{19}F chemical shifts of fluorinated RT and drug-resistant variants revealed that the fluorine resonance is a sensitive probe for identifying mutation-induced changes in the enzyme. Our data show that in the

unliganded enzyme, the NNRTI-binding pocket is highly plastic and not locked into a single conformation. Upon inhibitor binding, the binding pocket rigidifies.

In this thesis I show that solution NMR can be a powerful tool to study proteins that resist crystallization, and provide important unique information on protein structure and dynamics.

BIBLIOGRAPHY

1. (2015) HIV/AIDS. In: World Health Organization. www.who.int/mediacentre/factsheets/fs360/en/. Accessed 6 Mar 2016

2. Arts EJ, Hazuda DJ (2012) HIV-1 Antiretroviral Drug Therapy. Cold Spring Harb Perspect Med 2:7161–7161.

3. Looney D, Ma A, Johns S (2015) HIV therapy-the state of art. Curr Top Microbiol Immunol 389:1–29.

4. Le Grice SFJ (2012) Human Immunodeficiency Virus reverse transcriptase: 25 years of research, drug discovery, and promise. J Biol Chem 287:40850–40857.

5. Sluis-Cremer N (2014) The emerging profile of cross-resistance among the nonnucleoside HIV-1 reverse transcriptase inhibitors. Viruses 6:2960–2973.

6. Engelman A, Cherepanov P (2012) The structural biology of HIV-1: mechanistic and therapeutic insights. Nat Rev Micro 10:279–290.

7. Lucic B, Lusic M (2016) Connecting HIV-1 Integration and Transcription: a step forward towards new treatments. FEBS Lett 1–21.

8. Freed EO (2015) HIV-1 assembly, release and maturation. Nat Rev Micro 13:484–496.

9. Laskey SB, Siliciano RF A mechanistic theory to explain the efficacy of antiretroviral therapy. Nat Rev Micro 12:772–780.

10. Frankel AD, Young JA (1998) HIV-1: fifteen proteins and an RNA. Annu Rev Biochem 67:1–25.

11. Zhao G, Perilla JR, Yufenyuy EL, et al (2013) Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. Nature 497:643–646.

12. Craigie R, Bushman FD (2012) HIV DNA Integration. Cold Spring Harb Perspect Med 2:a006890.

13. Hare S, Gupta SS, Valkov E, et al (2010) Retroviral intasome assembly and inhibition of DNA strand transfer. Nature 464:232–236.

14. Vicenzi E, Poli G (2014) Human Retroviruses. Humana Press. Totowa, NJ. 1087:1–354.

15. Karn J, Stoltzfus CM (2012) Transcriptional and Posttranscriptional Regulation of HIV-1 Gene Expression. Cold Spring Harb Perspect Med 2:a006916.

16. Li L, Li HS, Pauza CD, et al (2005) Roles of HIV-1 auxiliary proteins in viral pathogenesis and host-pathogen interactions. Cell Res 15:923–934.

17. Strebel K (2013) HIV accessory proteins versus host restriction factors. Curr Opin Virol 3:692–699.

18. Geller R, Domingo-Calap P, Cuevas JM, et al (2015) The external domains of the HIV-1 envelope are a mutational cold spot. Nat Commun 6:8571.

19. Debouck C, Gorniak JG, Strickler JE, et al (1987) Human immunodeficiency virus protease expressed in Escherichia coli exhibits autoprocessing and specific maturation of the gag precursor. Proc Natl Acad Sci USA 84:8903–8906.

20. Hill M, Tachedjian G, Mak J (2005) The packaging and maturation of the HIV-1 Pol proteins. Curr HIV Res 3:73–85.

21. Karacostas V, Wolffe EJ, Nagashima K, et al (1993) Overexpression of the HIV-1 Gag-Pol Polyprotein Results in Intracellular Activation of HIV-1 Protease and Inhibition of Assembly and Budding of Virus-like Particles. Virology 193:661–671.

22. Miller M, Jaskolski M, Rao JKM, et al (1989) Crystal structure of a retroviral protease proves relationship to aspartic protease family. Nature 337:576–579.

23. Pettit SC, Lindquist JN, Kaplan AH, Swanstrom R (2005) Processing sites in the human immunodeficiency virus type 1 (HIV-1) gag-pro-pol precursor are cleaved by the viral protease at different rates. Retrovirology 2:66–6.

24. Haverkos HW, Curran JW (1982) The current outbreak of Kaposi's sarcoma and opportunistic infections. CA Cancer J Clin 32:330–339.

25. De Clercq E (1998) The role of non-nucleoside reverse transcriptase inhibitors (NNRTIs) in the therapy of HIV-1 infection. Antiviral Res 38:153–179.

26. Gulick RM, Mellors JW, Havlir D, et al (1997) Treatment with indinavir, zidovudine, and lamivudine in adults with Human Immunodeficiency Virus infection and prior antiretroviral therapy. N Engl J Med 337:734–739.

27. Staszewski S, Morales-Ramirez J (1999) Efavirenz plus zidovudine and lamivudine, efavirenz plus indinavir, and indinavir plus zidovudine and lamivudine in the treatment of HIV-1 infection in adults. N Engl J Med 341:1865–1873.

28. Robbins GK, De Gruttola V, Shafer RW (2003) Comparison of sequential three-drug regimens as initial therapy for HIV-1 infection. N Engl J Med 349:2293–2303.

29. Shafer RW, Smeaton LM, Robbins GK, et al (2003) Comparison of four-drug regimens and pairs of sequential three-drug regimens as initial therapy for HIV-1 infection. N Engl J Med 349:2304–2315.

30. Vella S, Schwartländer B, Sow SP, et al (2012) The history of antiretroviral therapy and of its implementation in resource-limited areas of the world. AIDS 26:1231–1241.

31. Roberts JD, Bebenek K, Kunkel TA (1988) The accuracy of reverse transcriptase from HIV-1. Science 242:1171–1173.

32. Wei X, Ghosh SK, Taylor ME, et al (1995) Viral dynamics in human immunodeficiency virus type 1 infection. Nature 373:117–122.

33. Santoro MM, Perno CF (2013) HIV-1 genetic variability and clinical implications. ISRN Microbiology 2013:1–20.

34. Keele BF, Giorgi EE, Salazar-Gonzalez JF, et al (2008) Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci USA 105:7552–7557.

35. Parrish NF, Gao F, Li H, et al (2013) Phenotypic properties of transmitted founder HIV-1. Proc Natl Acad Sci USA 110:6626–6633.

36. Kohlstaedt LA, Wang J, Friedman JM, Rice PA (1992) Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. Science 256:1783–1790.

37. Huang H (1998) Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. Science 282:1669–1675.

38. Hsiou Y, Ding J, Das K, et al (1996) Structure of unliganded HIV-1 reverse transcriptase at 2.7 Å resolution: implications of conformational changes for polymerization and inhibition mechanisms. Structure 4:853–860.

39. Jacobo-Molina A, Ding J, Nanni RG, et al (1993) Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA. Proc Natl Acad Sci USA 90:6320–6324.

40. Sarafianos SG, Clark AD, Das K, Tuske S (2002) Structures of HIV-1 reverse transcriptase with pre-and post-translocation AZTMP-terminated DNA. EMBO J 21:6614–6624.

41. Tuske S, Sarafianos SG, Clark AD, Ding J (2004) Structures of HIV-1 RT–DNA complexes before and after incorporation of the anti-AIDS drug tenofovir. Nat Struct Mol Biol 11:469–474.

42. Sarafianos SG, Das K, Tantillo C, et al (2001) Crystal structure of HIV-1 reverse transcriptase in complex with a polypurine tract RNA:DNA. EMBO J 20:1449–1461.

43. Sarafianos SG, Das K, Clark AD, et al (1999) Lamivudine (3TC) resistance in HIV-1 reverse transcriptase involves steric hindrance with beta-branched amino acids. Proc Natl Acad Sci USA 96:10027–10032.

44. Rodgers DW, Gamblin SJ, Harris BA, et al (1995) The structure of unliganded reverse transcriptase from the human immunodeficiency virus type 1. Proc Natl Acad Sci USA 92:1222–1226.

45. Ren J, Nichols CE, Stamp A, et al (2006) Structural insights into mechanisms of non-nucleoside drug resistance for HIV-1 reverse transcriptases mutated at codons 101 or 138. FEBS J 273:3850–3860.

46. Ren J, Milton J, Weaver KL, et al (2000) Structural basis for the resilience of efavirenz (DMP-266) to drug resistance mutations in HIV-1 reverse transcriptase. Structure 8:1089–1094.

47. Lapkouski M, Tian L, Miller JT, et al (2013) Complexes of HIV-1 RT, NNRTI and RNA/DNA hybrid reveal a structure compatible with RNA degradation. Nat Struct Mol Biol 20:230–236.

48. Lansdon EB, Samuel D, Lagpacan L, et al (2010) Visualizing the molecular interactions of a nucleotide analog, GS-9148, with HIV-1 reverse transcriptase-DNA complex. J Mol Biol 397:967–978.

49. Lansdon EB, Brendza KM, Hung M, et al (2010) Crystal structures of HIV-1 reverse transcriptase with etravirine (TMC125) and rilpivirine (TMC278): implications for drug design. J Med Chem 53:4295–4299.

50. Esnouf R, Ren J, Ross C, et al (1995) Mechanism of inhibition of HIV-1 reverse transcriptase by non-nucleoside inhibitors. Nat Struct Mol Biol 2:303–308.

51. Drozdzal P, Michalska K, Kierzek R, et al (2012) Structure of an RNA/DNA dodecamer corresponding to the HIV-1 polypurine tract at 1.6 Å resolution. Acta Crystallogr Sect D Biol Crystallogr 68:169–175.

52. Hsiou Y, Ding J, Das K, et al (1996) Structure of unliganded HIV-1 reverse transcriptase at 2.7 A resolution: implications of conformational changes for polymerization and inhibition mechanisms. Structure 4:853–860.

53. Ding J, Das K, Hsiou Y, et al (1998) Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody Fab fragment at 2.8 Å resolution. J Mol Biol 284:1095–1111.

54. Das K, Bandwar RP, White KL, et al (2009) Structural basis for the role of the K65R mutation in hiv-1 reverse transcriptase polymerization, excision antagonism, and tenofovir resistance. J Biol Chem 284:35092–35100.

55. Das K, Martinez SE, Bauman JD (2012) HIV-1 reverse transcriptase complex with DNA and nevirapine reveals non-nucleoside inhibition mechanism. Nat Struct Mol Biol 19:253–259.

56. Das K, Martinez SE, Bandwar RP, Arnold EE (2014) Structures of HIV-1 RT-RNA/DNA ternary complexes with dATP and nevirapine reveal conformational flexibility of RNA/DNA: insights into requirements for RNase H cleavage. Nucleic Acids Res 42:8125–8137.

57. Balzarini J, Das K, Bernatchez JA, et al (2015) Alpha-carboxy nucleoside phosphonates as universal nucleoside triphosphate mimics. Proc Natl Acad Sci USA 112:3475–3480.

58. Das K, Bauman JD, Clark AD Jr., et al (2008) High-resolution structures of HIV-1 reverse transcriptase/TMC278 complexes: strategic flexibility explains potency against resistance mutations. Proc Natl Acad Sci USA 105:1466–1471.

59. Zhan P, Chen X, Li D, et al (2013) HIV-1 NNRTIs: structural diversity, pharmacophore similarity, and implications for drug design. Med Res Rev 33 Suppl 1:E1–72.

60. Pauli W (1924) The theoretical significance of the satellites of some spectrum lines and the effect on them of magnetic fields. Naturwissenschaften 12:741.

61. Rabi II, Millman S, Kusch P, Zacharias JR (1939) The Molecular Beam Resonance Method for Measuring Nuclear Magnetic Moments. Phys Rev 55:526–535.

62. Saunders M, Wishnia A, Kirkwood JG (1957) The Nuclear Magnetic Resonance Spectrum Of Ribonuclease 1. J Am Chem Soc 79:3289–3290.

63. Bovey FA, Tiers GVD, Filipovich G (1959) Polymer NSR spectroscopy. I. The motion and configuration of polymer chains in solution. J Polym Sci A Polym Chem 38:73–90.

64. Cohen JS, Jaroszewski JW, Kaplan O (1995) A history of biological applications of NMR spectroscopy. Prog Nucl Magn Reson Spectrosc 28:53–98.

65. Emsley JW, Feeney J (1995) Milestones in the first fifty years of NMR. Prog Nucl Magn Reson Spectrosc 28:1–9.

66. Jardetzky O, Wade NG, Fischer JJ (1963) Proton magnetic resonance investigation of enzyme-coenzyme complexes. Nature 197:183–184.

67. Klein MP, Barton GW Jr (1963) Enhancement of signal-to-noise ratio by continuous averaging: application to magnetic resonance. Rev Sci Instrum 34:754–759.

68. Ernst RR, Anderson WA (1966) Application of fourier transform spectroscopy to magnetic resonance. Rev Sci Instrum 37:93–102.

69. McDonald CC, Phillips WD (1967) Manifestations of the tertiary structures of proteins in high-frequency nuclear magnetic resonance. J Am Chem Soc 89:6332–6341.

70. Aue WP, Bartholdi E, Ernst RR (1976) Two-dimensional spectroscopy. Application to nuclear magnetic resonance. J Chem Phys 64:2229–2246.

71. Kumar A, Wagner G, Ernst RR, Wuthrich K (1980) Studies of J-connectivities and selective ^{1}H-^{1}H Overhauser effects in H_2O solutions of biological macromolecules by two-dimensional NMR experiments. Biochem Biophys Res Commun 96:1156–1163.

72. Wagner G, Wüthrich K (1982) Amide proton exchange and surface conformation of the basic pancreatic trypsin inhibitor in solution. J Mol Biol 160:343–361.

73. Williamson MP, Havel TF, Wuthrich K (1985) Solution conformation of proteinase inhibitor IIA from bull seminal plasma by ^{1}H nuclear magnetic resonance and distance geometry. J Mol Biol 182:295–315.

74. Bax A (1994) Multidimensional nuclear magnetic resonance methods for protein studies. Curr Opin Struct Biol 4:738–744.

75. Clore GM, Gronenborn AM (1987) Determination of three-dimensional structures of proteins in solution by nuclear magnetic resonance spectroscopy. Protein Eng 1:275–288.

76. Clore GM, Nilges M, Sukumaran DK, et al (1986) The three-dimensional structure of alpha1-purothionin in solution: combined use of nuclear magnetic resonance, distance geometry and restrained molecular dynamics. EMBO J 5:2729–2735.

77. Griesinger C, Sorensen OW, Ernst RR (1987) Novel three-dimensional NMR techniques for studies of peptides and biological macromolecules. J Am Chem Soc 109:7227–7228.

78. Nilges M, Gronenborn AM, Brünger AT, Clore GM (1988) Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints. Application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. Protein Eng 2:27–38.

79. Nilges M, Clore GM, Gronenborn AM (1988) Determination of three-dimensional structures of proteins from interproton distance data by dynamical simulated annealing from a random array of atoms Circumventing problems associated with folding. FEBS Lett 239:129–136.

80. Oschkinat H, Griesinger C, Kraulis PJ, et al (1988) Three-dimensional NMR spectroscopy of a protein in solution. Nature 332:374–376.

81. Vuister GW, Boelens R, Kaptein R (1988) Nonselective three-dimensional NMR spectroscopy. The 3D NOE-HOHAHA experiment. J Magn Reson 80:176–185.

82. Fesik SW, Gampe RT, Zuiderweg ERP (1989) Heteronuclear three-dimensional NMR spectroscopy. Natural abundance ^{13}C chemical shift editing of ^{1}H-^{1}H COSY spectra. J Am Chem Soc 111:770–772.

83. Clore GM, Gronenborn AM (1994) Multidimensional heteronuclear nuclear magnetic resonance of proteins. Methods Enzymol 239:349–363.

84. Berman HM, Westbrook J, Feng Z, et al (2000) The Protein Data Bank. Nucleic Acids Res 28:235–242.

85. Wuthrich K (1990) Protein structure determination in solution by NMR spectroscopy. J Biol Chem 265:22059–22062.

86. Clore GM, Gronenborn AM (1998) Determining the structures of large proteins and protein complexes by NMR. Trends Biotechnol 16:22–34.

87. Wuthrich K (1998) The second decade--into the third millenium. Nat Struct Mol Biol 5 Suppl:492–495.

88. Cala O, Guilliere F, Krimm I (2014) NMR-based analysis of protein-ligand interactions. Anal Bioanal Chem 406:943–956.

89. Vinogradova O, Qin J (2012) NMR as a unique tool in assessment and complex determination of weak protein-protein interactions. Top Curr Chem 326:35–45.

90. Osawa M, Takeuchi K, Ueda T, et al (2012) Functional dynamics of proteins revealed by solution NMR. Curr Opin Struct Biol 22:660–669.

91. Fielding L (2007) NMR methods for the determination of protein–ligand dissociation constants. Prog Nucl Magn Reson Spectrosc 51:219–242.

92. Sekhar A, Kay LE (2013) NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. Proc Natl Acad Sci USA 110:12867–12874.

93. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. J Magn Reson 160:65–73.

94. Marion D (2013) An Introduction to Biological NMR Spectroscopy. Mol Cell Proteomics 12:3006–3025.

95. Sattler M, Schleucher J (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution. Prog Nucl Magn Reson Spectrosc 34:93–158.

96. Grzesiek S, Bax A (1992) An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. J Magn Reson 99:201–207.

97. Grzesiek S, Bax A (1993) Amino acid type determination in the sequential assignment procedure of uniformly $^{13}C/^{15}N$-enriched proteins. J Biomol NMR 3:185–204.

98. Bax A, Clore GM, Gronenborn AM (1990) ^{1}H-^{1}H correlation via isotropic mixing of ^{13}C magnetization, a new three-dimensional approach for assigning ^{1}H and ^{13}C spectra of ^{13}C-enriched proteins. J Magn Reson 88:425–431.

99. Kumar A, Ernst RR, Wuthrich K (1980) A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. Biochem Biophys Res Commun 95:1–6.

100. Ramachandran GN, Sasisekharan V (1968) Conformation of polypeptides and proteins. Adv Protein Chem 23:283–438.

101. Hafsa NE, Arndt D, Wishart DS (2015) CSI 3.0: a web server for identifying secondary and super-secondary structure in proteins using NMR chemical shifts. Nucleic Acids Res 43:W370–W377.

102. Shen Y, Delaglio F, Cornilescu G, Bax A (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. J Biomol NMR 44:213–223.

103. Tjandra N (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline Medium. Science 278:1111–1114.

104. Hansen MR, Hanson P, Pardi A (2000) Filamentous bacteriophage for aligning RNA, DNA, and proteins for measurement of nuclear magnetic resonance dipolar coupling interactions. In: RNA - Ligand Interactions, Part A. Elsevier, pp 220–240

105. Bax A, Grishaev A (2005) Weak alignment NMR: a hawk-eyed view of biomolecular structure. Curr Opin Struct Biol 15:563–570.

106. Iwahara J, Anderson DE, Murphy EC, Clore GM (2003) EDTA-derivatized deoxythymidine as a tool for rapid determination of protein binding polarity to DNA by intermolecular paramagnetic relaxation enhancement. J Am Chem Soc 125:6634–6635.

107. Battiste JL, Wagner G (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. Biochemistry-Us 39:5355–5365.

108. Iwahara J, Tang C, Marius Clore G (2007) Practical aspects of ^{1}H transverse paramagnetic relaxation enhancement measurements on macromolecules. J Magn Reson 184:185–195.

109. Schmitz C, John M, Park AY, et al (2006) Efficient χ-tensor determination and NH assignment of paramagnetic proteins. J Biomol NMR 35:79–87.

110. Clore GM, Iwahara J (2009) Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. Chem Rev 109:4108–4139.

111. Riek R, Pervushin K, Wuthrich K (2000) TROSY and CRINEPT: NMR with large molecular and supramolecular structures in solution. Trends Biochem Sci 25:462–468.

112. Rule GS, Hitchens TK (2006) Fundamentals of protein NMR spectroscopy. Springer Ccience & Business media.

113. Wuthrich K, Wider G (2003) Transverse relaxation-optimized NMR spectroscopy with biomacromolecular structures in solution. Magn Reson Chem 41:S80–S88.

114. Pervushin K, Riek R, Wider G, Wüthrich K (1997) Attenuated T2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proc Natl Acad Sci USA 94:12366–12371.

115. Gardner KH, Kay LE (1998) The use of ^{2}H, ^{13}C, ^{15}N multidimensional NMR to study the structure and dynamics of proteins. Annu Rev Biophys Biomol Struct 27:357–406.

116. Salzmann M, Pervushin K, Wider G (2000) NMR assignment and secondary structure determination of an octameric 110 kDa protein using TROSY in triple resonance experiments. J. Am. Chem. Soc.

117. Gelis I, Bonvin AMJJ, Keramisanou D, et al (2007) Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR. Cell 131:756–769.

118. Sprangers R, Kay LE (2007) Quantitative dynamics and binding studies of the 20S proteasome by NMR. Nature 445:618–622.

119. Kerfah R, Plevin MJ, Sounier R, et al (2015) Methyl-specific isotopic labeling: a molecular tool box for solution NMR studies of large proteins. Curr Opin Struct Biol 32:113–122.

120. Plevin MJ, Boisbouvier J (2012) Isotope-labelling of methyl groups forNMR studies of large proteins. In: Clore M, Potts J (eds) Recent Developments in Biomolecular NMR. The Royal Society of Chemistry, pp 1–24

121. Hajduk PJ, Augeri DJ, Mack J, Mendoza R (2000) NMR-Based Screening of Proteins Containing ^{13}C-Labeled Methyl Groups. J Am Chem Soc 122:7898–7904.

122. Goto NK, Gardner KH, Mueller GA, et al (1999) A robust and cost-effective method for the production of Val, Leu, Ile (δ1) methyl-protonated ^{15}N-, ^{13}C-, ^{2}H-labeled proteins. 13:369–374.

123. Gross JD, Gelev VM, Wagner G (2003) A sensitive and robust method for obtaining intermolecular NOEs between side chains in large protein complexes. 25:235–242.

124. Gardner KH, Kay LE (1997) Production and incorporation of ^{15}N, ^{13}C, ^{2}H (^{1}H-δ1 methyl) isoleucine into proteins for multidimensional NMR studies. J Am Chem Soc 119:7599–7600.

125. Fischer M, Kloiber K, Häusler J, et al (2007) Synthesis of a ^{13}C-Methyl-Group-Labeled Methionine Precursor as a Useful Tool for Simplifying Protein Structural Analysis by NMR Spectroscopy. ChemBioChem 8:610–612.

126. Ayala I, Sounier R, Usé N, et al (2009) An efficient protocol for the complete incorporation of methyl-protonated alanine in perdeuterated protein. J Biomol NMR 43:111–119.

127. Godoy-Ruiz R, Guo C, Tugarinov V (2010) Alanine methyl groups as NMR probes of molecular structure and dynamics in high-molecular-weight proteins. J Am Chem Soc 132:18340–18350.

128. Isaacson RL, Simpson PJ, Liu M, et al (2007) A new labeling method for methyl transverse relaxation-optimized spectroscopy NMR spectra of alanine residues. J Am Chem Soc 129:15428–15429.

129. Danielson MA, Falke JJ (1996) Use of ^{19}F NMR to probe protein structure and conformational changes. Annu Rev Biophys Biomol Struct 25:163–195.

130. Gerig JT (1994) Fluorine NMR of proteins. Prog Nucl Magn Reson Spectrosc 26:293–370.

131. Kitevski-Leblanc JL, Prosser RS (2012) Current applications of ^{19}F NMR to studies of protein structure and dynamics. Prog Nucl Magn Reson Spectrosc 62:1–33.

132. Campos-Olivas R, Aziz R, Helms GL, et al (2002) Placement of 19F into the center of GB1: effects on structure and stability. FEBS Lett 517:55–60.

133. Arseniev AS, Kuryatov AB, Tsetlin VI, Bystrov VF (1987) ^{19}F NMR study of 5-fluorotryptophan-labeled bacteriorhodopsin. FEBS Lett 213:283–288.

134. Chaiken IM, Freedman MH, Lyerla JRJ, Cohen JS (1973) Preparation and studies of ^{19}F-labeled and enriched ^{13}C-labeled semisynthetic ribonuclease-S' analogues. J Biol Chem 248:884–891.

135. Didenko T, Liu JJ, Horst R, et al (2013) Fluorine-^{19} NMR of integral membrane proteins illustrated with studies of GPCRs. Curr Opin Struct Biol 23:740–747.

136. Marsh ENG, Suzuki Y (2014) Using ^{19}F NMR to probe biological interactions of proteins and peptides. ACS Chem Biol 9:1242–1250.

137. Liu JJ, Horst R, Katritch V, et al (2012) Biased signaling pathways in β_2-adrenergic receptor characterized by ^{19}F-NMR. Science 335:1106–1110.

138. Rydzik AM, Brem J, van Berkel SS, et al (2014) Monitoring conformational changes in the NDM-1 metallo-β-lactamase by 19F NMR spectroscopy. Angew Chem Int Ed 53:3129–3133.

139. Luck LA, Falke JJ (1991) ^{19}F NMR studies of the D-galactose chemosensory receptor. 1. Sugar binding yields a global structural change. Biochemistry-Us 30:4248–4256.

140. Kitevski-Leblanc JL, Hoang J, Thach W, et al (2013) 19F NMR Studies of a Desolvated Near-Native Protein Folding Intermediate. Biochemistry-Us 52:5780–5789.

141. Khan F, Kuprov I, Craggs TD, et al (2006) ^{19}F NMR studies of the native and denatured states of green fluorescent Protein. J Am Chem Soc 128:10729–10737.

142. Li C, Lutz EA, Slade KM, et al (2009) ^{19}F NMR studies of α-synuclein conformation and fibrillation. Biochemistry-Us 48:8578–8584.

143. Suzuki Y, Brender JR, Hartman K, et al (2012) Alternative pathways of human islet amyloid polypeptide aggregation distinguished by ^{19}F nuclear magnetic resonance-detected kinetics of monomer consumption. Biochemistry-Us 51:8154–8162.

144. Fischer M, Schott A-K, Kemter K, et al (2003) Riboflavin synthase of Schizosaccharomyces pombe. Protein dynamics revealed by ^{19}F NMR protein perturbation experiments. BMC Biochem 4:1–18.

145. Sykes BD, Hull WE (1978) Fluorine nuclear magnetic resonance studies of proteins. Methods Enzymol 49:270–295.

146. Cellitti SE, Jones DH, Lagpacan L, et al (2008) In vivoIncorporation of Unnatural Amino Acids to Probe Structure, Dynamics, and Ligand Binding in a Large Protein by Nuclear Magnetic Resonance Spectroscopy. J Am Chem Soc 130:9268–9281.

147. Peeler JC, Mehl RA (2011) Site-specific incorporation of unnatural amino acids as probes for protein conformational changes. Methods Mol Biol 794:125–134.

148. Frieden C, Hoeltzli SD, Bann JG (2004) The preparation of 19F-labeled proteins for NMR studies. Methods Enzymol 380:400–415.

149. Kim HW, Perez JA, Ferguson SJ, Campbell ID (1990) The specific incorporation of labelled aromatic amino acids into proteins through growth of bacteria in the presence of glyphosate. Application to fluorotryptophan labelling to the H(+)-ATPase of Escherichia coli and NMR studies. FEBS Lett 272:34–36.

150. Kitevski-Leblanc JL, Al-Abdul-Wahid MS, Prosser RS (2009) A mutagenesis-free approach to assignment of ^{19}F NMR resonances in biosynthetically labeled proteins. J Am Chem Soc 131:2054–2055.

151. Noren CJ, Anthony-Cahill SJ, Griffith MC, Schultz PG (1989) A general-method for site-specific incorporation of unnatural amino-acids into proteins. Science 244:182–188.

152. Studier FW (2005) Protein production by auto-induction in high density shaking cultures. PREP 41:207–234.

153. Sharaf NG, Gronenborn AM (2015) ^{19}F-modified proteins and ^{19}F-containing ligands as tools in solution NMR studies of protein interactions. Methods Enzymol 565:67–95.

154. Eickbush TH, Jamburuthugoda VK (2008) The diversity of retrotransposons and the properties of their reverse transcriptases. Virus Res 134:221–234.

155. Wilhelm M, Wilhelm FX (2001) Reverse transcription of retroviruses and LTR retrotransposons. Cell Mol Life Sci 58:1246–1262.

156. Temin HM (1993) Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. Proc Natl Acad Sci USA 90:6900–6903.

157. Temin HM, Mizutani S (1992) RNA-dependent DNA polymerase in virions of Rous sarcoma virus. 1970. Biotechnology 24:51–56.

158. Baltimore D (1970) Viral RNA-dependent DNA Polymerase: RNA-dependent DNA Polymerase in Virions of RNA Tumour Viruses. Nature 226:1209–1211.

159. Hizi A, Herschhorn A (2008) Retroviral reverse transcriptases (other than those of HIV-1 and murine leukemia virus): a comparison of their molecular and biochemical properties. Virus Res 134:203–220.

160. Coffin JM, Hughes SH, Varmus HE (1997) Retroviruses. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY)

161. Katz RA, Skalka AM (1994) The retroviral enzymes. Annu Rev Biochem 63:133–173.

162. Herschhorn A, Hizi A (2010) Retroviral reverse transcriptases. Cell Mol Life Sci 67:2717–2747.

163. Venezia CF, Meany BJ, Braz VA, Barkley MD (2009) Kinetics of Association and Dissociation of HIV-1 Reverse Transcriptase Subunits. Biochemistry-Us 48:9084–9093.

164. Sluis-Cremer N, Arion D, Abram ME, Parniak MA (2004) Proteolytic processing of an HIV-1 pol polyprotein precursor: insights into the mechanism of reverse transcriptase p66/p51 heterodimer formation. Int J Biochem Cell Biol 36:1836–1847.

165. Divita G, Rittinger K, Geourjon C, et al (1995) Dimerization kinetics of HIV-1 and HIV-2 reverse transcriptase: a two step process. J Mol Biol 245:508–521.

166. Sharma SK, Fan N, Evans DB (1994) Human immunodeficiency virus type 1 (HIV-1) recombinant reverse transcriptase. Asymmetry in p66 subunits of the p66/p66 homodimer. FEBS Lett 343:125–130.

167. Chattopadhyay D, Evans DB, Deibel MR, et al (1992) Purification and characterization of heterodimeric human immunodeficiency virus type 1 (HIV-1) reverse transcriptase produced by in vitro processing of p66 with recombinant HIV-1 protease. J Biol Chem 267:14227–14232.

168. Wang J, Smerdon SJ, Jäger J, et al (1994) Structural basis of asymmetry in the human immunodeficiency virus type 1 reverse transcriptase heterodimer. Proc Natl Acad Sci USA 91:7242–7246.

169. Davies JF, Hostomska Z, Hostomsky Z, et al (1991) Crystal structure of the ribonuclease H domain of HIV-1 reverse transcriptase. Science 252:88–95.

170. Jacobo-Molina A, Arnold EE (1991) Hiv Reverse-Transcriptase Structure-Function-Relationships. Biochemistry-Us 30:6351–6361.

171. Hostomska Z, Matthews DA, Davies JF, et al (1991) Proteolytic release and crystallization of the RNase H domain of human immunodeficiency virus type 1 reverse transcriptase. J Biol Chem 266:14697–14702.

172. Geoffrey A Mueller, Koteppa Pari, Eugene F DeRose, et al (2004) Backbone Dynamics of the RNase H Domain of HIV-1 Reverse Transcriptase. Biochemistry-Us 43:9332–9342.

173. Powers R, Clore GM, Stahl SJ, Wingfield PT (1992) Analysis of the backbone dynamics of the ribonuclease H domain of the human immunodeficiency virus reverse transcriptase using ^{15}N relaxation measurements. Biochemistry-Us 31:9150–9157.

174. Tatyana I Igumenova, Kendra King Frederick A, Wand AJ (2006) Characterization of the Fast Dynamics of Protein Amino Acid Side Chains Using NMR Relaxation in Solution. Chem Rev 106:1672–1699.

175. Palmer AG (1997) Probing molecular motion by NMR. Curr Opin Struct Biol 7:732–737.

176. Kay LE (2005) NMR studies of protein structure and dynamics. J Magn Reson 173:193–207.

177. Zheng X, Mueller GA, Derose EF, London RE (2009) Solution characterization of [methyl-^{13}C]methionine HIV-1 reverse transcriptase by NMR spectroscopy. Antiviral Res 84:205–214.

178. Zheng X, Mueller GA, Cuneo MJ, et al (2010) Homodimerization of the p51 Subunit of HIV-1 Reverse Transcriptase. Biochemistry-Us 49:2821–2833.

179. Schneider A, Peter D, Schmitt J, Leo B (2014) Structural requirements for enzymatic activities of foamy virus protease-reverse transcriptase. Proteins: Structure. 82:375-3785.

180. Kay LE, Gardner KH (1997) Solution NMR spectroscopy beyond 25 kDa. Curr Opin Struct Biol 7:722–731.

181. Wishart DS, Sykes BD, Richards FM (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. Biochemistry-Us 31:1647–1651.

182. Fletcher RS, Holleschak G, Nagy E, et al (1996) Single-step purification of recombinant wild-type and mutant HIV-1 Reverse Transcriptase. PREP 7:27–32.

183. Restle T, Müller B, Goody RS (1990) Dimerization of human immunodeficiency virus type 1 reverse transcriptase. A target for chemotherapeutic intervention. J Biol Chem 265:8986–8988.

184. Marko RA, Liu H-W, Ablenas CJ, et al (2013) Binding kinetics and affinities of heterodimeric versus homodimeric HIV-1 reverse transcriptase on DNA-DNA substrates at the single-molecule level. J Phys Chem B 117:4560–4567.

185. Weigelt J (1998) Single scan, sensitivity-and gradient-enhanced TROSY for multidimensional NMR experiments. J Am Chem Soc 120:10778–10779.

186. Vranken WF, Boucher W, Stevens TJ, et al (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins 59:687–696.

187. Delaglio F, Grzesiek S, Vuister GW, et al (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6:277–293.

188. Wishart DS, Sykes BD (1994) The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. J Biomol NMR 4:171–180.

189. Shen Y, Bax A (2007) Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. J Biomol NMR 38:289–302.

190. Pervushin K (2000) Impact of transverse relaxation optimized spectroscopy (TROSY) on NMR as a technique in structural biology. Q Rev Biophys 33:161–197.

191. Schreiber G, Haran G, Zhou HX (2009) Fundamental aspects of protein-protein association kinetics. Chem Rev 109:839–860.

192. Schlosshauer M, Baker D (2004) Realistic protein-protein association rates from a simple diffusional model neglecting long-range interactions, free energy barriers, and landscape ruggedness. Protein Science 13:1660–1669.

193. Pari K, Mueller GA, DeRose EF, Kirby TW (2003) Solution Structure of the RNase H Domain of the HIV-1 Reverse Transcriptase in the Presence of Magnesium. Biochemistry-Us 42:639–650.

194. Christen MT, Menon L, Myshakina NS, et al (2012) Structural Basis of the Allosteric Inhibitor Interaction on the HIV-1 Reverse Transcriptase RNase H Domain. Chem Biol Drug Des 80:706–716.

195. Jäger J, Smerdon SJ, Wang J, et al (1994) Comparison of three different crystal forms shows HIV-1 reverse transcriptase displays an internal swivel motion. Structure 2:869–876.

196. Tomasselli AG, Sarcich JL, Barrett LJ, et al (1993) Human immunodeficiency virus type-1 reverse transcriptase and ribonuclease H as substrates of the viral protease. Protein Sci 2:2167–2176.

197. Zheng X, Pedersen LC, Gabel SA, et al (2014) Selective unfolding of one Ribonuclease H domain of HIV reverse transcriptase is linked to homodimer formation. Nucleic Acids Res 42:5361–5377.

198. Das K, Lewi PJ, Hughes SH, Arnold EE (2005) Crystallography and the design of anti-AIDS drugs: conformational flexibility and positional adaptability are important in the design of non-nucleoside HIV-1 reverse transcriptase inhibitors. Prog Biophys Mol Biol 88:209–231.

199. Sarafianos SG, Marchand B, Das K, et al (2009) Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. J Mol Biol 385:693–713.

200. De Béthune M-P (2010) Non-nucleoside reverse transcriptase inhibitors (NNRTIs), their discovery, development, and use in the treatment of HIV-1 infection: A review of the last 20 years (1989–2009). Antiviral Res 85:75–90.

201. Sluis-Cremer N, Temiz NA, Bahar I (2004) Conformational changes in HIV-1 reverse transcriptase induced by nonnucleoside reverse transcriptase inhibitor binding. Curr HIV Res 2:323–332.

202. Wapling J, Moore KL, Sonza S, Mak J (2005) Mutations That Abrogate Human Immunodeficiency Virus Type 1 Reverse Transcriptase Dimerization Affect Maturation of the Reverse Transcriptase Heterodimer. J Virol 79:10247–10257.

203. Braz VA, Holladay LA, Barkley MD (2010) Efavirenz binding to HIV-1 reverse transcriptase monomers and dimers. Biochemistry-Us 49:601–610.

204. Cabodevilla JF, Odriozola L, Santiago E, Martinez-Irujo JJ (2001) Factors affecting the dimerization of the p66 form of HIV-1 reverse transcriptase. Eur J Biochem 268:1163–1172.

205. Beard WA, Wilson SH (1993) Kinetic analysis of template:primer interactions with recombinant forms of HIV-1 reverse transcriptase. Biochemistry-Us 32:9745–9753.

206. Divita G, Rittinger K, Geourjon C, et al (1995) Dimerization kinetics of HIV-1 and HIV-2 reverse transcriptase: A two step process. J Mol Biol 245:508–521.

207. Zheng X, Perera L, Mueller GA, et al (2015) Asymmetric conformational maturation of HIV-1 reverse transcriptase. Elife 4:11952.

208. Sharaf NG, Poliner E, Slack RL, et al (2014) The p66 immature precursor of HIV-1 reverse transcriptase. Proteins 82:2343–2352.

209. Fletcher RS, Holleschak G, Nagy E, et al (1996) Single-step purification of recombinant wild-type and mutant HIV-1 reverse transcriptase. PREP 7:27–32.

210. Venezia CF, Howard KJ, Ignatov ME, Holladay LA (2006) Effects of efavirenz binding on the subunit equilibria of HIV-1 reverse transcriptase. Biochemistry-Us 45:2779–2789.

211. Sattler M, Maurer M, Schleucher J (1995) A simultaneous $^{15}N,^{1}H$- and $^{13}C,^{1}H$-HSQC with sensitivity enhancement and a heteronuclear gradient echo. 5:97–102.

212. Vranken WF, Boucher W, Stevens TJ, et al (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins 59:687–696.

213. Schwieters CD, Kuszewski JJ, Clore GM (2006) Using Xplor–NIH for NMR molecular structure determination. Prog Nucl Magn Reson Spectrosc 48:47–62.

214. Cornilescu G, Delaglio F, Bax A (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. 13:289–302.

215. Laskowski RA, Rullmann JAC, MacArthur MW, et al (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8:477–486.

216. Davis IW, Leaver-Fay A, Chen VB, et al (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 35:W375–W383.

217. Koradi R, Billeter M, Wuthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. J Mol Gr 14:51–55.

218. Humphrey W, Dalke A, Schulten K (1996) VMD: Visual molecular dynamics. J Mol Gr 14:33–38.

219. Pettersen EF, Goddard TD, Huang CC, et al (2004) UCSF Chimera—A visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612.

220. Schrödinger, LLC (2015) The PyMOL molecular graphics system, version 1.8.

221. Clark SA, Tronrud DE, Karplus PA (2015) Residue-level global and local ensemble-ensemble comparisons of protein domains. Protein Science 24:1528–1542.

222. Lapkouski M, Tian L, Miller JT, et al (2013) Complexes of HIV-1 RT, NNRTI and RNA/DNA hybrid reveal a structure compatible with RNA degradation. Nat Struct Mol Biol 20:230–236.

223. Powell MD, Beard WA, Bebenek K, et al (1999) Residues in the αH and αI helices of the HIV-1 reverse transcriptase thumb subdomain required for the specificity of RNase H-catalyzed removal of the polypurine tract primer. J Mol Biol 274:19885–19893.

224. Betancor G, Puertas MC, Nevot M, et al (2010) Mechanisms involved in the selection of HIV-1 reverse transcriptase thumb subdomain polymorphisms associated with nucleoside analogue therapy failure. Antimicrob Agents Chemother 54:4799–4811.

225. Bebenek K, Beard WA, Casas-Finet JR, et al (1995) Reduced frameshift fidelity and processivity of HIV-1 reverse transcriptase mutants containing alanine substitutions in helix H of the thumb subdomain. J Biol Chem 270:19516–19523.

226. Beard WA, Stahl SJ, Kim HR, Bebenek K (1994) Structure/function studies of human immunodeficiency virus type 1 reverse transcriptase. Alanine scanning mutagenesis of an alpha-helix in the thumb subdomain. J Biol Chem 269:28091–28097.

227. Hermann T, Meier T, Götte M, Heumann H (1994) The “helix clamp” in HIV-1 reverse transcriptase: a new nucleic acid binding motif common in nucleic acid polymerases. Nucleic Acids Res 22:4625–4633.

228. Tu X, Das K, Han Q, et al (2010) Structural basis of HIV-1 resistance to AZT excision. Nat Struct Mol Biol 17:1202-1209.

229. Pauwels R (2004) New non-nucleoside reverse transcriptase inhibitors (NNRTIs) in development for the treatment of HIV infections. Curr Opin Pharmacol 4:437–446.

230. Kensch O, Restle T, Wöhrl BM, et al (2000) Temperature-dependent equilibrium between the open and closed conformation of the p66 subunit of HIV-1 reverse transcriptase revealed by site-directed spin labelling. J Mol Biol 301:1029–1039.

231. Seckler JM, Barkley MD, Wintrode PL (2011) Allosteric suppression of HIV-1 reverse transcriptase structural dynamics upon Inhibitor Binding. Biophys J 100:144–153.

232. Zhou Z, Madrid M, Evanseck JD, Madura JD (2005) Effect of a bound non-nucleoside RT inhibitor on the dynamics of wild-type and mutant HIV-1 reverse transcriptase. J Am Chem Soc 127:17253–17260.

233. Ivetac A, McCammon JA (2009) Elucidating the inhibition mechanism of HIV-1 non-nucleoside reverse transcriptase inhibitors through multicopy molecular dynamics simulations. J Mol Biol 388:644–658.

234. Shen L, Shen J, Luo X, et al (2003) Steered molecular dynamics simulation on the binding of NNRTI to HIV-1 RT. Biophys J 84:3547–3563.

235. Bahar I, Erman B, Jernigan RL, et al (1999) Collective motions in HIV-1 reverse transcriptase: examination of flexibility and enzyme function. J Mol Biol 285:1023–1037.

236. Seckler JM, Leioatts N, Miao H, Grossfield A (2013) The interplay of structure and dynamics: insights from a survey of HIV-1 reverse transcriptase crystal structures. Proteins 81:1792–1801.

237. Monroe JI, El-Nahal WG, Shirts MR (2014) Investigating the mutation resistance of nonnucleoside inhibitors of HIV-RT using multiple microsecond atomistic simulations. Proteins 82:130–144.

238. Rodríguez-Barrios F, Gago F (2004) Understanding the basis of resistance in the irksome Lys103Asn HIV-1 reverse transcriptase mutant through targeted molecular dynamics simulations. J Am Chem Soc 126:15386–15387.

239. Wright DW, Hall BA, Kellam P, Coveney PV (2012) Global conformational dynamics of HIV-1 reverse transcriptase bound to non-nucleoside inhibitors. Biology 1:222–244.

240. Das K, Ding J, Hsiou Y, et al (1996) Crystal Structures of 8-Cl and 9-Cl TIBO Complexed with Wild-type HIV-1 RT and 8-Cl TIBO Complexed with the Tyr181Cys HIV-1 RT Drug-resistant Mutant. J Mol Biol 264:1085–1100.

241. Richman DD, Havlir D, Corbeil J, et al (1994) Nevirapine resistance mutations of human immunodeficiency virus type 1 selected during therapy. J Virol 68:1660–1666.

242. Reuman EC, Rhee S-Y, Holmes SP, Shafer RW (2010) Constrained patterns of covariation and clustering of HIV-1 non-nucleoside reverse transcriptase inhibitor resistance mutations. J Antimicrob Chemother 65:1477–1485.

243. Bacheler L, Jeffrey S, Hanna G, et al (2001) Genotypic correlates of phenotypic resistance to efavirenz in virus isolates from patients failing nonnucleoside reverse transcriptase inhibitor therapy. J Virol 75:4999–5008.

244. Rhee S-Y, Liu T, Ravela J, et al (2004) Distribution of human immunodeficiency virus type 1 protease and reverse transcriptase mutation patterns in 4,183 persons undergoing genotypic resistance testing. Antimicrob Agents Chemother 48:3122–3126.

245. Tambuyzer L, Vingerhoets J, Azijn H, et al (2010) Characterization of genotypic and phenotypic changes in HIV-1-infected patients with virologic failure on an etravirine-containing regimen in the DUET-1 and DUET-2 clinical studies. AIDS Res Hum Retroviruses 26:1197–1205.

246. Tambuyzer L, Nijs S, Daems B, et al (2011) Effect of mutations at position E138 in HIV-1 reverse transcriptase on phenotypic susceptibility and virologic response to etravirine. J Acquir Immune Defic Syndr 58:18–22.

247. Rimsky L, Vingerhoets J, Van Eygen V, et al (2012) Genotypic and phenotypic characterization of HIV-1 isolates obtained from patients on rilpivirine therapy experiencing virologic failure in the phase 3 ECHO and THRIVE studies: 48-week analysis. J Acquir Immune Defic Syndr 59:39–46.

248. Azijn H, Tirry I, Vingerhoets J, et al (2010) TMC278, a next-generation nonnucleoside reverse transcriptase inhibitor (NNRTI), active against wild-type and NNRTI-resistant HIV-1. Antimicrob Agents Chemother 54:718–727.

249. Zhang Z, Xu W, Koh Y-H, et al (2007) A novel nonnucleoside analogue that inhibits human immunodeficiency virus type 1 isolates resistant to current nonnucleoside reverse transcriptase inhibitors. Antimicrob Agents Chemother 51:429–437.

250. Hsiou Y, Ding J, Das K, et al (2001) The Lys103Asn mutation of HIV-1 RT: a novel mechanism of drug resistance. J Mol Biol 309:437–445.

251. Geretti AM, Easterbrook P (2001) Antiretroviral resistance in clinical practice. Int J STD AIDS 12:145–153.